Faculty of Technology
Centre for Computational Intelligence, School of Computer Science and Informatics

---

Application of Computational Intelligence in Cognitive Radio Network for Efficient Spectrum Utilization, and Speech Therapy

PhD THESIS

Sunday Iliya

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

August, 2016

I dedicate this thesis to God the Father, the Son and the Holy Spirit, and to my wife Rose Sunday Iliya

Abstract

With spectrum becoming an ever scarcer resource, it is critical that new communication systems utilize all the available frequency bands as efficiently as possible in time, frequency and spatial domains. Society requires more high capacity and broadband wireless connectivity, demanding greater access to spectrum. Most of the licensed spectrums are grossly underutilized while some spectrum (licensed and unlicensed) are overcrowded. The problem of spectrum scarcity and underutilization can be minimized by adopting a new paradigm of wireless communication scheme. Advanced CR network or Dynamic Adaptive Spectrum Sharing is one of the ways to optimize our wireless communications technologies for high data rates while maintaining users' desired quality of service (QoS) requirements. Scanning a wideband spectrum to find spectrum holes to deliver to users an acceptable quality of service using algorithmic methods requires a lot of time and energy. Computational Intelligence (CI) techniques can be applied to these scenarios to predict the available spectrum holes, and the expected RF power in the channels. This will enable the CR to predictively avoid noisy channels among the idle channels, thus delivering optimum QoS at less radio resources.

In this study, spectrum holes search using artificial neural network (ANN) and traditional search methods were simulated. The RF power traffic of some selected channels ranging from 50MHz to 2.5GHz were modelled using optimized ANN and support vector machine (SVM) regression models for prediction of real world RF power. The prediction accuracy and generalization was improved by combining different prediction models with a weighted output to form one model. The meta-parameters of the prediction models were evolved using population based differential evolution and swarm intelligence optimization algorithms.

The success of CR network is largely dependent on the overall world knowledge of spectrum utilization in both time, frequency and spatial domains. To identify underutilized bands that can serve as potential candidate bands to be exploited by CRs, spectrum occupancy survey based on long time RF measurement using energy detector was conducted. Results show that the average spectrum utilization of the bands considered within the studied location is less than 30%.

Though this research is focused on the application of CI with CR as the main target, the skills and knowledge acquired from the PhD research in CI was applied in some neighbourhood areas related to the medical field. This includes the use of ANN and SVM for impaired speech segmentation which is the first phase of a research project that aims at developing an artificial speech therapist for speech impaired patients.

Declaration

This is to certify that:

(i) the thesis comprises only my original work towards the PhD except where indicated,

(ii) due acknowledgement has been made in the text to all other material used,

(iii) the thesis is less than 40,000 words in length, exclusive of table, maps, bibliographies, appendices and footnotes.

Sunday Iliya

Publications

During the course of this project, a number of publications and public presentations have been made based on the work presented in this thesis. They are listed here for reference.

Journals

- **Amos Olusegun Abioye, Adeola Kola-Mustapha, George Tangyie Chi and Sunday Iliya** *Quantification of in situ granulation-induced changes in pre-compression, solubility, dose distribution and intrinsic in vitro release characteristics of ibuprofen–cationic dextran conjugate crystanules*, International Journal of Pharmaceutics, (2014). Impact Factor: 4.09.
- **Sunday Iliya and Ferrante Neri** *Towards artificial speech therapy: A neural system for impaired speech segmentation*, International Journal of Neural Systems (IJNS), 17 March, 2016. Impact Factor: 6.507.
- **Sunday Iliya, Eric Goodyer, John Gow, Jethro Shell and Mario Gongora** *Application of Artificial Neural Network and Support Vector Regression in Cognitive Radio Networks for RF Power Prediction Using Compact Differential Evolution Algorithm*, Annals of computer science and information systems, (2015).
- **Sunday Iliya, Eric Goodyer, John Gow, Mario Gongora and Jethro Shell** *Spectrum occupancy survey, Leicester, UK, for cognitive radio application*, International journal of scientific and engineering research (IJSER), (2015). Impact factor 3.8
- **Sunday Iliya, Eric Goodyer, John Gow, Mario Gongora and Jethro Shell** *Spectrum hole prediction and white space ranking using artificial neural network for cognitive radio application*, International journal of scientific and technology research (IJST), (2015). Impact factor 3.023.

Submitted Journal

- **Sunday Iliya, Eric Goodyer, John Gow, Jethro Shell and Mario Gongora** *Towards spectrum efficient utilization: Spectrum occupancy survey and a neural system for RF power prediction*, IEEE Transactions on Neural Networks and Learning Systems (IEEETNNLS), (2016). Submitted and awaiting acceptance. Impact Factor: 4.291.

Conference Papers

- **Arinze Akutekwe, Sunday Iliya and Huseyin Seker** *An Optimized Hybrid Dynamic Bayesian Network Approach using Differential Evolution Algorithm for Hepatocellular Carcinoma Diagnosis* IEEE International Conference on Adaptive Science and Technology (ICAST), (2014).

- **Sunday Iliya, Eric Goodyer, Jethro Shell, Mario Gongora and John Gow** *Optimized Neural Network Using Differential Evolutionary and Swarm Intelligence Optimization Algorithms for RF Power Prediction in Cognitive Radio Network: A Comparative Study*,IEEE International Conference on Adaptive Science and Technology (ICAST), (2014).

- **Sunday Iliya, Eric Goodyer, Mario Gongora and Jethro Shell and John Gow** *Optimized Artificial Neural Network Using Differential Evolution for Prediction of RF Power in VHF/UHF TV and GSM 900 Bands for Cognitive Radio Networks* IEEE United Kingdom Computational Intelligence (UKCI) Conference, (2014).

- **Sunday Iliya, Ferrante Neri, Dylan Menzies, Pip Cornelius and Lorenzo Picinali** *"Differential Evolution Schemes for Speech Segmentation: A Comparative Study*, IEEE symposia series on computational intelligence (SSCI), (2014).

- **Sunday Iliya, Dylan Menzies, Ferrante Neri, Pip Cornelius and Lorenzo Picinali** *Robust Impaired Speech Segmentation Using Neural Network Mixture Model*, IEEE Symposium on Signal Processing and Information Technology (ISSPIT), (2014).

- **Sunday Iliya, Eric Goodyer, John Gow, Jethro Shell and Mario Gongora** *Application of Artificial Neural Network and Support Vector Regression in Cognitive Radio Networks for RF Power Prediction Using Compact Differential Evolution Algorithm*, IEEE Federated Conference on Computer Science and Information Systems (FedCSIS), (2015).

Acknowledgements

First of all, I would like to express my gratitude to God Almighty who in his infinite mercy have seen me through to the end of my PhD, may his name be praised forever. I sincerely thank my supervisor in the person of Professor Eric Goodyer who out of his busy schedules has contributed immensely to the success of my PhD. I want to appreciate his encouragement, support, advice and prompt response to my needs within the period of my study, may God remember him and his family for good. My gratitude also goes to the other supervisory team members for their assistance and support. I want to knowledge the contribution of Dr. Daniel Paluszczyszyn for volunteering to read the thesis, and for other commitments toward the success of the entire research.

I would like to express my sincere gratitude to my fellow PhD colleagues and members of staff who has contributed to the success of my academic career. I assured you that all your contributions are greatly appreciated.

This acknowledgement cannot be complete without acknowledging the assistance of my wife and our son Gurumnen for their support and encouragement. May God bless them and be gracious unto them. I also want to thank our extended family members for their payers and support, and the members of the Redeemed Christian Church of God (RCCG), green pastures parish Leicester, and Members of Anglican communion in Nigeria for their incredible contribution towards the success of my study.

The speech therapy research presented in Chapter 5 was supported by the Higher Education Innovation Fund at De Montfort University, UK. I wish to thank Dr. Dylan Menzies for his valuable suggestions, Dr. Lorenzo Picinali for having supervised the speech recording used in this study and Mrs. Pip Cornelius for having provided me with her expertise in speech therapy.

Finally, I sincerely appreciate my sponsor i.e. petroleum technology development fund (PTDF) scholarship board for sponsoring my PhD. May God richly bless Nigeria and remember the country for good. I appreciate the commitment of University of Agriculture Makurdi, Nigeria, for approving my study leave even while I'm already in the UK. I also want to thank De Montfort University and DIGITs for the opportunity to serve in various capacities such as front runner position for music and gesture recognition, assistant research fellow for artificial speech therapist research, research fellow position for developing water advisory and evaluation system, and also part time lectureship position for three academic sessions. May God richly bless, and remember Great Britain for good and restore all its glory. Amen

Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1

Introduction

1.1 Introduction

Due to the current static spectrum allocation policy, most of the licensed radio spectra are not maximally utilized and often free (idle) while some spectra are overcrowded. Hence, the current spectrum scarcity is the direct consequence of the static spectrum allocation policy and not the fundamental lack of spectrum. The first bands to be approved for Cognitive Radio (CR) communication by the US Federal Communication Commission (FCC) because of their gross underutilization in time, frequency and spatial domain are the very high frequency (VHF) and ultra high frequency (UHF) TV bands Communication (2003), Valenta et al. (2010), Haykin et al. (2009), Mitola and Maguire (1999), Bhattacharya et al. (2011). Though spectrum activities varies with time, frequency and spatial domain; this research explored only two varying domains i.e. time and frequency while the spatial domain is kept constant as the experiments were conducted in one geographical location within the same altitude. This research is focus on the study of real world RF power distribution of some selected channels, and the development of RF power prediction models for dynamic spectrum access as shown in Figure 1.3. The channels considered in this study range from 50MHz to 860MHz, consisting of radio and TV broadcasting channels within the VHF/UHF bands; and 868MHz to 2.5GHz which include ISM 868 band, GSM900 and GSM1800 bands and ISM 2.4GHz band. Scanning a very wide band in search of spectrum holes (white spaces) using algorithmic or stochastic approaches requires a lot of time and energy. To circumvent this problem, computational intelligence based RF power predictive models are developed. This will help the cognitive radio (CR) to avoid noisy channels and also to prioritise its scanning by starting with the channels that are predicted to be free. The channel with the highest probability of being idle (least predicted RF power) are scanned first followed by the second and so forth. The problem of spectrum scarcity and underutilization can be minimized by adopting a new paradigm of wireless communication scheme. Advanced Cognitive Radio (CR) network or Adaptive Spectrum Sharing (ASS) is one of the ways to optimize our wireless communications technologies for high data rates in a dynamic environment while maintaining users' desired quality of service (QoS) requirements. CR is a radio equipped with the capability of awareness,

perception, adaptation and learning of its radio frequency (RF) environment Oh and Choi (2009). CR is an intelligent radio where much of the signal conditioning and digital signal processing that was traditionally done in static hardware are implemented via software. Irrespective of the definition of CR, it has the followings basic features: observation, adaptability and intelligence as shown in Figure 1.1. CR radio as coined by Joseph Mitola is meant to operate at a higher level of intelligence and adaptability as compared with software defined radio Mitola and Maguire (1999), Adane et al. (2009; 2008). CR is the key enabling tool for dynamic spectrum access and a promising solution for the present problem of spectrum scarcity and underutilization. Cognitive radio network is made up of two users i.e. the license owners called the primary users (PU) who are the incumbent legitimate owners of the spectrum and the cognitive radios commonly called the secondary users (SU) who intelligently and opportunistically access the unused licensed spectrum based on some agreed conditions. CR access to licensed spectrum is subject to two constraints i.e. on a no interference base, this implies that CRs can use the licensed spectrum only when the license owners are not using the channel. This is referred to as the overlay cognitive radio scheme. The second constraint is on the transmitted power, in this case, SU (CR) can coexist with the PU as long as the aggregate interference at the PU receivers is below a given threshold which will not be harmful to the PU nor degrade the QoS requirements of the PU, this is referred to as the underlay CR network scheme Stevenson et al. (2009), Communication (2003). The maximum permissible interference limit in underlay CR scheme is called the interference temperature limit. There are four major steps involved in cognitive radio network, these are: spectrum sensing, spectrum decision, spectrum sharing, and spectrum mobility Xing et al. (2013), Wyglinski et al. (2009).

In spectrum sensing, the CR senses the PU Rf activity using one or more of the following spectrum sensing algorithms such as energy detector, cyclostationary features detector, cooperative sensing, match filter detector, eigenvalue detector, etc. to sense the occupancy status of the PU. Based on the sensing results, the CR will take a decision using a binary classifier to classify the PU channels (spectrum) as either busy or idle thereby identifying the white spaces (spectrum holes or idle channels). Spectrum sharing deals with efficient allocation of the available white spaces to the CRs within a given geographical location at a given period of time while spectrum mobility is the ability of the CR to vacate the channels when the PU reclaimed ownership of the channel and search for another spectrum hole to communicate. During the withdrawal or search period, the CR should maintain seamless communication. The cognitive cycle can be summarised as shown in Figure 1.2. Many wireless broadband devices ranging from simple communication to complex systems automation, are deployed daily with increasing demand for more, this calls for optimum utilization of the limited spectrum resources via CR paradigm. Future wireless communication devices should be enhanced with cognitive capability for optimum spectrum utilization.

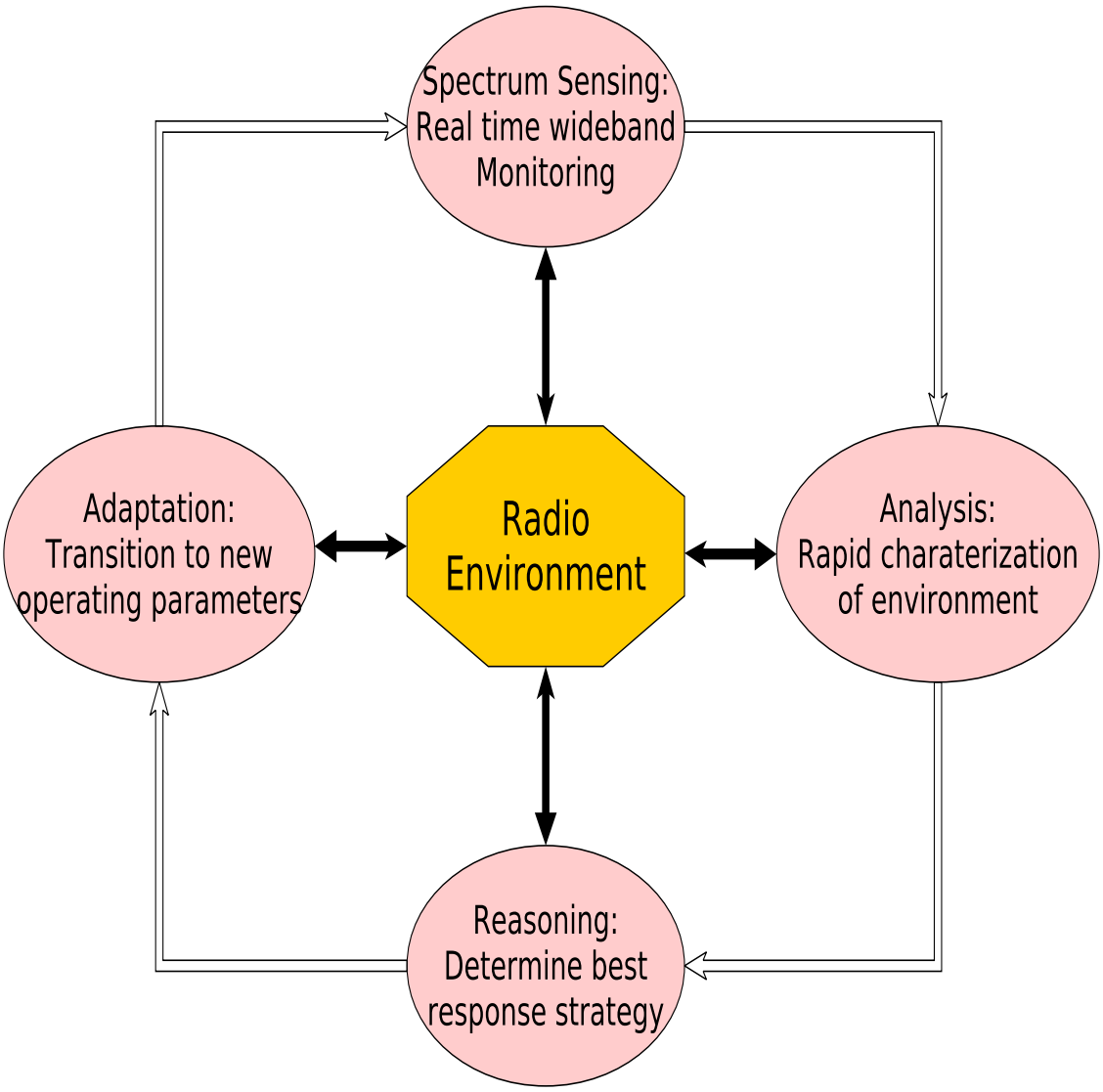

Figure 1.1: Cognitive radio capabilities

1.1.1 Aim and Objectives

The main aim of this research is to develop spectrum holes and RF power predictive models using computational intelligence for cognitive radio application. The second objective is to conduct an intensive spectrum occupancy survey, and to estimate the spectrum occupancy level of the studied location (i.e. DMU, Leicester) thereby identifying underutilised bands that can serve as potential candidate bands to be used for CR application. Impaired speech segmentation using algorithmic signal processing, and CI techniques were implemented, towards the realization of stand alone artificial speech therapist.

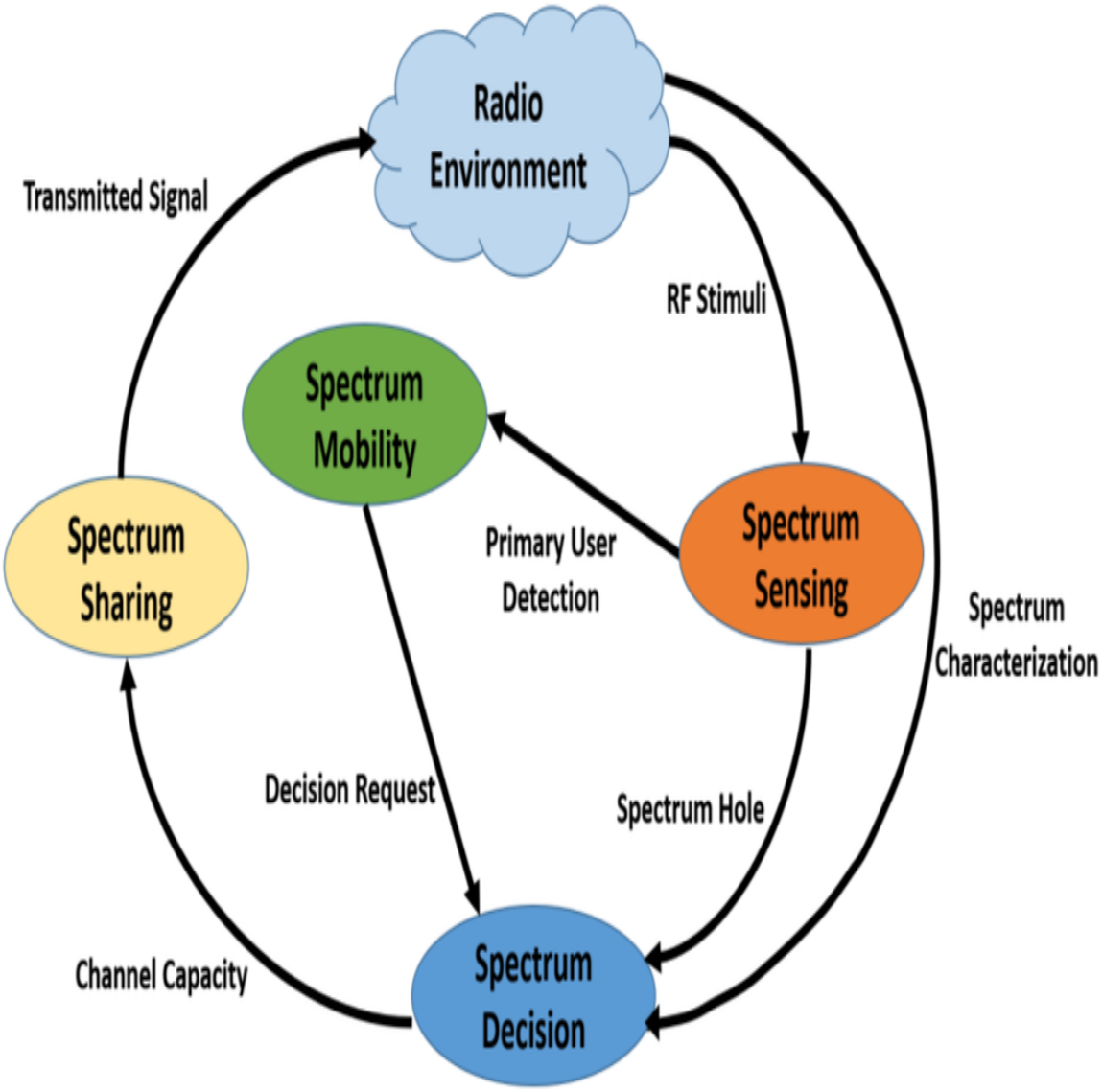

Figure 1.2: Cognitive cycle

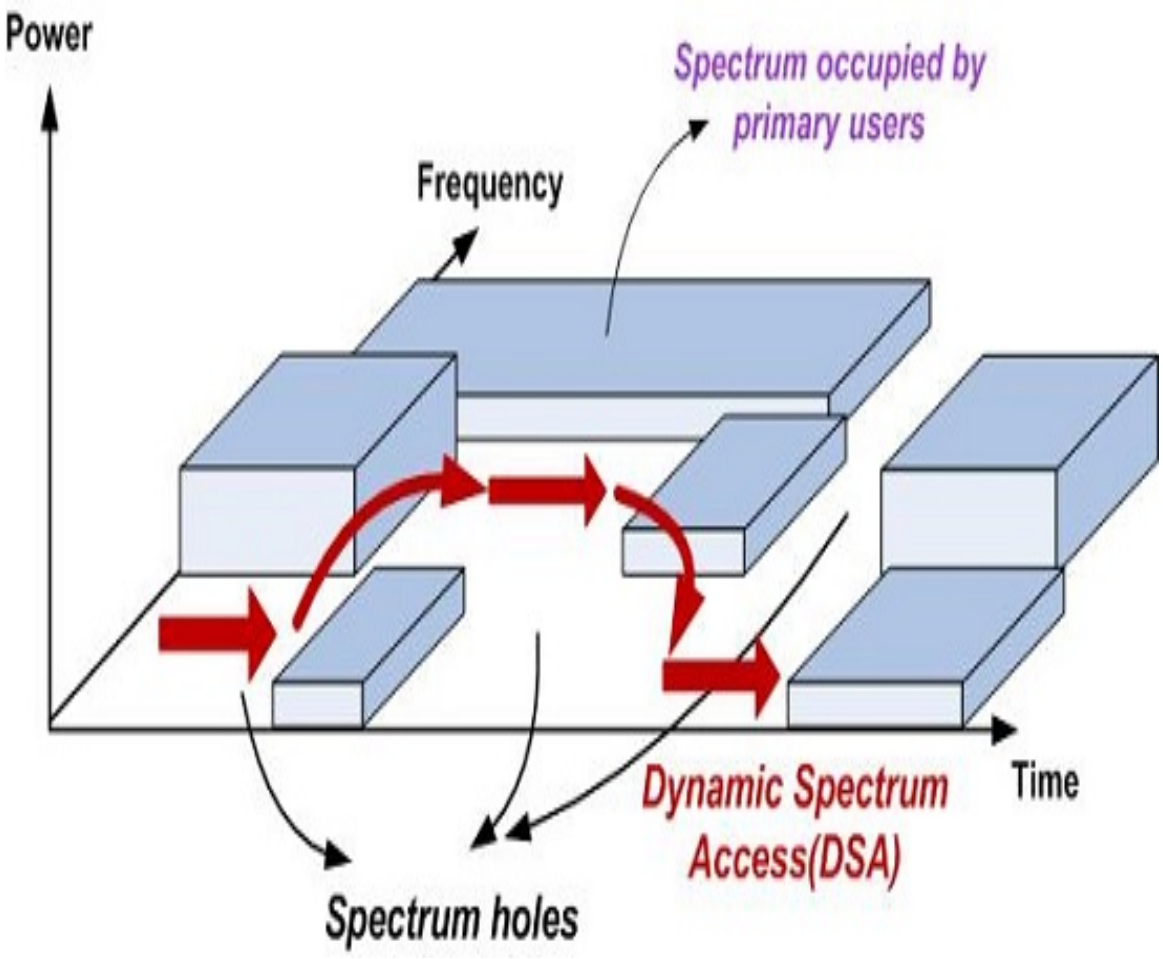

Figure 1.3: Dynamic spectrum access

1.2 Contribution of the thesis

The contributions of this study are listed as follows:

- Spectrum occupancy or utilization estimate of some communication bands in Leicester, UK was conducted. These bands include: 50MHz to 860MHz consisting of radio and TV broadcasting channels; and 868MHz to 2.5GHz with emphases on ISM 868 band, GSM900 and GSM1800 bands and ISM 2.4GHz band, some bands within these ranges where also considered. The survey aim at identifying underutilised bands that can serve as potential candidate bands to be exploited for CR application. It is quite interesting to note that the outcome of the survey reveals that the average utilization estimate of the bands within the studied location is less than 30% of the entire bands considered.

- Optimised ANN and SVR models for prediction of real world RF power traffic of the studied location were developed. These models are used for prediction of RF power and instantaneous occupancy status of the channels. This will enable the CR to avoid noisy channels among the predicted idle channels for effective communication at less radio resources such as bandwidth and transmitted power. The input of the RF power prediction models does not contain any RF power related parameters such as signal to noise ratio or a priory knowledge of primary users' signals, thus making the models robust for CR application. Spectrum hole (idle channel) prediction using probabilistic ANN and algorithmic approach were developed and compared using simulated data. The word probabilistic here referred to the fact that the outputs of the ANN gives the probability of the channels been idle.

- The thesis proposes a robust and novel way of obtaining a benchmark for the weights of a combined prediction model that captures the advantages of equal weight average method, inverse ranking method and inverse mean square error method of evolving weights of a combined forecasting model as explained in Section 4.4. The proposed hybrid benchmark tuned using differential evolution variants, proof to converge quickly to a good solution as shown in Figure 4.10. Six prediction models consisting of different topologies of ANN and SVR models are combined together with weighted outputs to form a single prediction model. As a proof of concept, the weights of the combined model are evolved using the proposed method and the overall prediction accuracy was found to increase with an average of 46% as compare with the individual models used.

- The work presented in Cherkassky and Ma (2004) was improved by using differential evolutionary and particle swarm optimization algorithms to evolves the meta parameters of the SVR models which was found to increase the accuracy and generalization of the models. The SVR models were trained by the proposed nested ad-hoc hybrid algorithmic framework consisting of metaheuristics optimization algorithm variants and convex optimization algorithm. The models

used implemented a novel and innovative initial weight optimization of the ANN models. The initial weights of the ANN are evolved using deferential evolutionary and particle swarm optimization algorithms. These population based randomization algorithms are used as global searchers which provides effective means of minimising premature convergence at local optimum during the initial training phase of the ANN. The weights are finally fine tuned towards the global optimum using single solution back propagation algorithms (BPA) at low learning rate and momentum constant to serve as a local searcher.

- The skills and knowledge acquired from this research were extended to develop computational intelligence models for application in medical field. Impaired speech segmentation models were developed by aggregating algorithmic speech signal processing approach, and CI techniques. The segmentation is the first phase of a research that aims at developing an artificial speech therapist for speech impaired patients.

1.3 Thesis Structure

Chapter 2 contains a literature review and background information needed for understanding of the subsequent chapters. Spectrum sensing techniques and the metaheuristics optimization algorithms used to optimise the prediction models are discussed. The theory of ANN and SVM used in modelling the radio frequency (RF) power prediction models are presented in this chapter. The metaparameters of the prediction models are optimised using hybrid nested algorithm consisting of differential evolution and swarm intelligence optimization algorithm variants, and convex optimization algorithm.

Chapter 3 forms one of the contributions of this thesis where long time spectrum occupancy survey using real world RF data captured at different period of time within three years was conducted. A brief description of the services within the bands considered is also highlighted. This is followed by the methods used in estimating the utilization of the bands. The chapter is concluded with the results of the survey, and brief summary of the findings.

Chapter 4 detailed the implementation of the RF power prediction models, and the topologies adopted. Sensitivity analysis is used to determine how many past RF power samples should be used as part of the input of the forecasting models for reliable current RF power prediction. The accuracy and generalization of the prediction models are improved by combining different models together to form one prediction model. There are two major contributions of this chapter, these are: The training of the SVM regression model using a hybrid nested ad-hoc algorithm and the training of the ANN using hybrid algorithm consisting of population based metaheuristics as global searchers and single solution back propagation algorithm as local searcher. The second contribution is a novel benchmark proposed for evolving the weight of a combined forecasting model. The chapter is concluded with the summary of the prediction results.

Chapter 5 consists of one major areas related to medical field where the knowledge and experiences acquired from the PhD research in CI is applied. The chapter starts with an introduction that gives a brief summary of the areas where CI has been applied to solve some problems related to medical field, and the outcome of each. This include the application of CI for impaired speech segmentation which is the first phase of the research that aims at developing an artificial speech therapist capable of providing audiovisual feedback progress report to speech impaired patient. Thus serving as a self training device in addition to the support received from human therapist. The rest of the chapter is dedicated to the detailed explanations of the methodology employed in designing the CI models for the impaired speech segmentation. This is because all the signal processing algorithms, and the CI models used for the impaired speech segmentation are derivatives of the outcome of the PhD research. The chapter is concluded with the results of the impaired speech segmentation and future work towards the realization of a standalone artificial speech therapist.

Chapter 6 contains the summary of the findings of this research and future work towards improving on the work presented in this thesis.

Chapter 2

Literature Review

2.1 Introduction

The thesis presents an efficient way of spectrum resource utilization using predictive models developed by aggregating different computational intelligence (CI) and machine learning techniques for cognitive radio (CR) application. This chapter contains the background information needed for understanding of the subsequent chapters and is organised as follows: One of the key enabling tools for cognitive radio network implementation is spectrum sensing algorithms which are presented in Section 2.2. Section 2.2.1 details spectrum sensing using energy detector which is the algorithm used in this thesis for estimating spectrum utilization with respect to the studied location. A brief review of some areas where CI techniques are used for cognitive radio application are discussed in Section 2.3. The general overview of evolutionary algorithms is presented in Section 2.4.1. The radio frequency (RF) power prediction models developed, consist of artificial neural network (ANN) and support vector machine (SVM) presented in Sections 2.5.1 and 2.5.2 respectively. The metaheuristic optimization algorithms used to optimised the predictive models are depicted in Section 2.6. The metaparameters of the prediction models are optimised using differential evolution and swarm intelligence optimization algorithms as detailed in Sections 2.6.1 and 2.6.5 respectively, while Section 2.8 gives the summary of the findings.

2.2 Spectrum Sensing Methods

For CR to exploit primary user (PU) spectrum, the CR has to sense the occupancy status of the licensed spectrum. Spectrum sensing aimed at determining spectrum holes and the presence of license owners. Spectrum sensing as applied to CR refer to a general term that involves obtaining the spectrum usage characteristics across multiple domains such as time, space, frequency, and code. This research explored only two varying domains i.e. time and frequency while the spatial domain is kept constant as the experiments were conducted in one geographical location within the same al-

titude. There are four major functional blocks in cognitive radio network, these are: spectrum sensing, spectrum decision or spectrum management, spectrum sharing, and spectrum mobility. Spectrum sensing is currently believed to be the most important or crucial component of cognitive radio network. This is because the success or performance of all the other three units are absolutely dependent on the accuracy and reliability of the spectrum sensing results, Subhedar and Birajdar (2011b), Yucek and Arslan (2009). There are two ongoing sensing operations to be carried out by the CR, the first is called out-of-band spectrum sensing in which the CR searches a wide band of frequency in search of spectrum holes to use, while the second phase commonly referred to as in-band sensing in which the CR keeps on monitoring the PU activity during the transmission period to vacate the channel thereby avoiding interference when the PU returns to active state, Lee (2009). There are many spectrum sensing methods that has been proposed, each with it merits and demerits. Spectrum sensing algorithms can be classified into two broad groups, these are blind and non-blind methods, Ekram et al. (2009), Wyglinski et al. (2010), Subhedar and Birajdar (2011b), Garhwal and Bhattacharya (2011). The blind algorithms does not required a priory knowledge of primary users signal structure neither made any assumption about the PU signals e.g. energy detection method is a blind spectrum sensing algorithms. While the non-blind algorithms required a prior knowledge of the incumbent signal characteristics for detection, among these algorithms are: cyclostationary feature detection, match filter detection, covariance based detection, waveform based detection, etc. All the spectrum sensing techniques can be modelled as a binary hypothesis testing methods. The two hypothesis to be tested are H_0 and H_1 given by Equation (2.1).

$$\begin{aligned} H_0 : &\quad y(n) = w(n) \\ H_1 : &\quad y(n) = hx(n) + w(n) \end{aligned} \tag{2.1}$$

where $w(n)$ is additive white Gaussian noise, $x(n)$ is PU signal at time n, h is the channel gain, $y(n)$ is the measured signal while H_0 and H_1 are the two hypothesis for absent and present of PU respectively.

Spectrum sensing techniques can also be classified based on the need or reason for engaging in spectrum sensing. These includes spectrum sensing for spectrum opportunity and spectrum sensing for interference detection. Spectrum sensing for spectrum opportunity is meant to be achieved by means of primary transmitter detection and cooperative detection. In primary transmitter detection, the detection of PUs is done by detecting the PUs transmitter signal type or signal strength. The detection methods for achieving this type of detection include matched filter (MF) based detection, energy detection, cyclostationary features detection, covariance detection, waveform detection, radio identification detection and random hough transform detection, Ekram et al. (2009), Wyglinski et al. (2010), Subhedar and Birajdar (2011b), Kamil and Yuan (2010). In cooperative detection, spectrum sensing is carried out collaboratively by the CRs in cooperative way depending on the method adopted as detailed in Section 2.2.4.

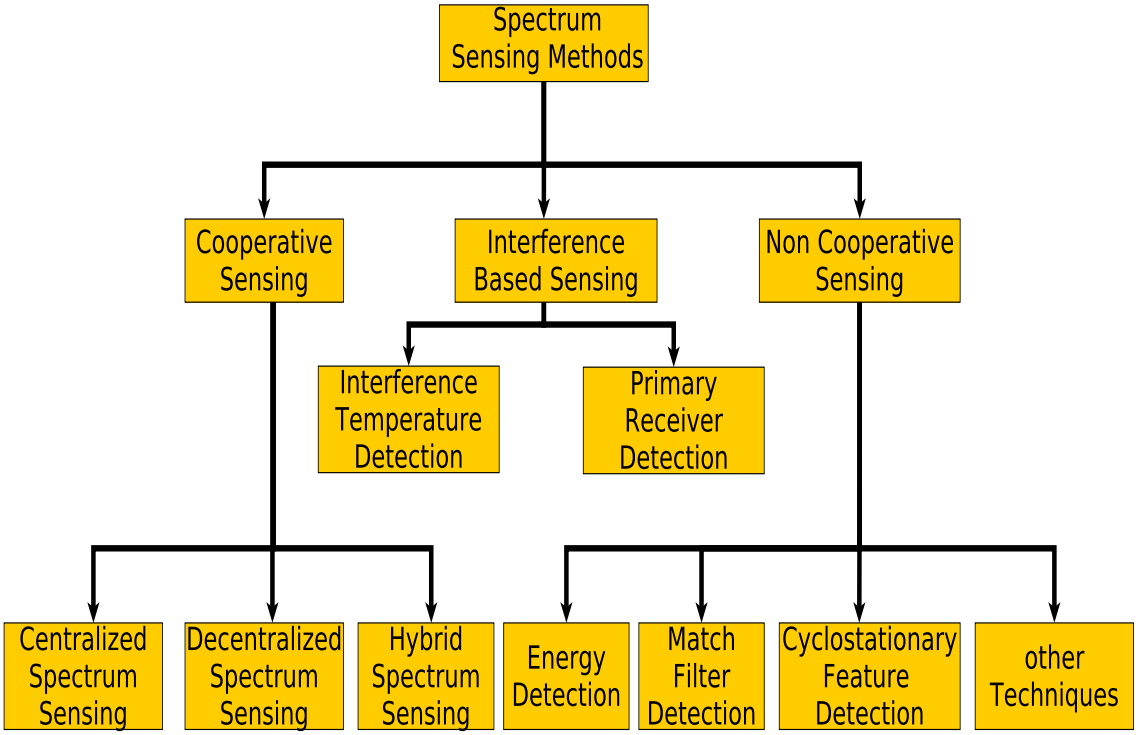

Figure 2.1: Spectrum sensing methods

2.2.1 Energy detection

PU has the flexibility and agility for chosen their modulation type and pulse shaping schemes which may not be known by the secondary user (SU) or CR. In situations where PU signal structure is unknown to the SU, the energy detector (ED) is the optimal, Ekram et al. (2009), Wyglinski et al. (2010), Subhedar and Birajdar (2011b), Kamil and Yuan (2010). In ED spectrum sensing technique, the power of the signal in a given channel is measured, this is compared with a predefined threshold for decision making. ED is the test of the following two binary hypothesis shown in Equation (2.2). The test statistics of ED is given by Equation (2.3). The time domain ED consist of a channel filter to reject out of band noise and adjacent signals, a Nyquist sampling analog to digital (A/D) converter, a square law and an averaging unit as shown in Figure 2.2. ED can also be implemented in frequency domain by averaging the frequency bins of the fast furrier transform (FFT) as depicted in Figure 2.3. The major advantages of ED lies in the simplicity of the algorithm and the fact that it does not required any a priori knowledge of the PU signal type, rather it only require the knowledge of the noise parameter. The main disadvantage of energy detection is that it is not robust where the noise variance is not know or is time variant which is often the case. Also the ED detector performance is very poor when the signal to noise ratio (SNR) is very low, and can not differentiate between PU signal and SU signal or noise since the PU signal is modelled as a zero-mean stationary white Gaussian process. In ED, if the test statistics is greater than the set threshold, the PU is assumed to be present otherwise

the channel is vacant. The choice of the threshold depends on the relative cost of false alarm and miss detection. The energy detector is used in this research for spectrum occupancy survey discussed in Chapter 3.

$$\begin{aligned} 0: &\quad y(n) = w(n) \\ 1: &\quad y(n) = hx(n) + w(n) \end{aligned} \tag{2.2}$$

where $w(n)$ is the additive white Gaussian noise, $x(n)$ is PU signal at time n, h is the channel gain, $y(n)$ is the measured signal while 0 and 1 are the two hypothesis for absent and present of PU respectively.

$$T = \sum_{k=1}^{N} (y(k))^2 \tag{2.3}$$

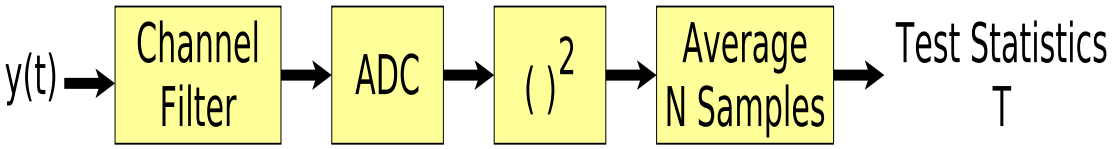

Figure 2.2: Prefiltered square law energy detector

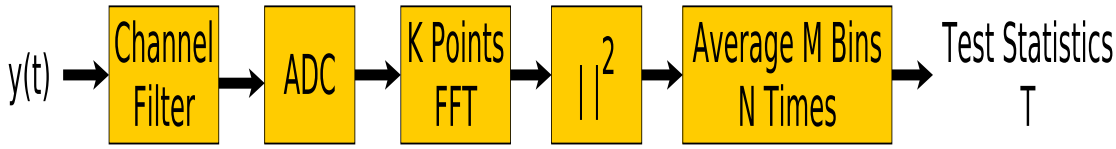

Figure 2.3: Energy detector using Welch's periodogram

2.2.2 Cyclostationary features detector

The sinusoidal carriers, pulse trains, hopping sequences, spreading codes, cyclic prefixes of modulated PU signals are periodic. This periodicity of cyclostationary signals is exploited by cyclostationary features detector (CFD) for detection of the present of PU. These signals exhibit features of spectral correlation and periodic statistics which distinguished them from noise and interferences. A signal is said to exhibit cyclostationarity if and only if the signal is correlated with certain frequency shifted-versions of itself, Gardner (1991), Garhwal and Bhattacharya (2011). The cyclic spectral density (CSD) gives peak values when the cyclic frequency is equal to the fundamental frequencies of the transmitted signals. The main advantage of CFD is that it can distinguish the PUs signals from noise and can detect PU signals with low signals to noise ratio. It main demerit is that the algorithm is complex with long observation or sensing time. It also require a priory knowledge of PU signal which may not be available to CRs. The

received signal is giving by Equation (2.4) while the cyclic spectral density function is as shown in Equation (2.5), Gardner (1991).

$$y(n) = hx(n) + w(n) \tag{2.4}$$

Where $w(n)$ is the additive white Gaussian noise, $x(n)$ is PU signal at time n, h is the channel gain, $y(n)$ is the measured signal

$$S(f, \alpha) = \sum_{\tau=-\infty}^{\infty} R_y^{\alpha}(\tau) e^{-j2\pi ft} \tag{2.5}$$

Where $R_y^{\alpha}(\tau) = E[y(n+\tau)y_*(n-\tau)e^{j2\pi fn}]$ is cyclic autocorrelation function, f and α are the fundamental signal frequency and cyclic frequency respectively and τ is time shift.

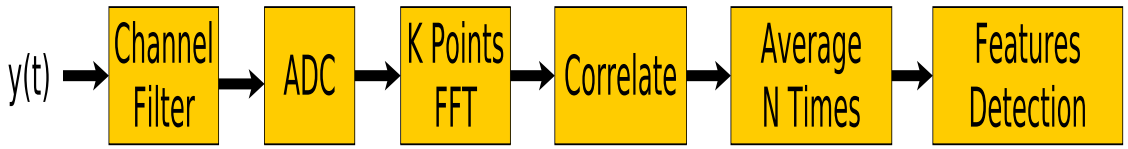

Figure 2.4: Cyclostationar feature detector

2.2.3 Match filter detection

When the secondary user has perfect knowledge of the PU signals, match filter (MF) is the optimum detector in stationary Gaussian noise. MF is a linear filter designed with special characteristics such that the signal to noise ratio (SNR) at the output of the filter for a given input signal is maximized, Proakis and Salehi (2008). MF operation is similar to correlation in which the PU or unknown signal is convolved with a filter whose impulse response is the mirror and time-shifted version of the reference signal for maximizing the output SNR. In MF the input is passed through a bandpass filter (BPF), after which it is convolved with the match filter whose impulse response is the same as the reference signal. The output of the match filter is then compared with a threshold for decision making. Detection using MF is very fast because it required only few samples to meet up with a given probability of detection constrain. At low SNR, the number of samples required is of the order $1/SNR$ in contrast to energy detector which the required number of samples is of the other $1/SNR^2$, Wyglinski et al. (2010). However, MF demodulate PU signals, hence CR require perfect knowledge of PU signal. Also it performance is poor in the event of frequency or time offset. The complexity of MF is further increase since CR need to have a dedicated receiver for all types of PU signals for effective detection. MF is not efficient in terms of power or energy consumption as various receiver algorithms need to be run for all types of PU signals. The block diagram of MF is as shown in Figure 2.5 while the MF detector is

modelled as:

$$y(n) = \sum_{k=-\infty}^{\infty} h(n-k)x(k) \tag{2.6}$$

Where x is the unknown signal vector that is convolved with the impulse response h of the filter and y is the output.

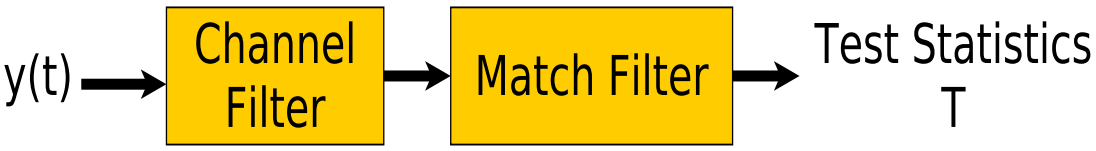

Figure 2.5: Match filter detector

2.2.4 Cooperative sensing

In cooperative sensing, the spectrum sensing task is carried out by multiple CRs working collaboratively. This approach aims at solving spectrum sensing problem resulting from noise uncertainty, fading, shadowing and faulty sensor of a single detector, Yucek and Arslan (2009). The reliability of detecting weak primary signal using one CR is difficult to maintain due to poor channel conditions such as multipath fading and shadowing between the primary users and the secondary users. Users at different distant geographical locations experience independent fading condition, the effects of destructive channel conditions resulting from fading, shadowing or hidden nodes can be mitigated if the users operate cooperatively. This will increase the overall detection reliability. When cooperation is allowed among users, individual secondary users required less sensitivity to achieve high detection reliability Zarrin (2011). Cooperative sensing was found to decrease significantly both the probability of false alarm and miss detection, and at the same time offering more protection to primary users, it also decreases the sensing time as compared with single local sensing methods. Cooperative sensing make use of the advantages of spatial diversities of CRs to improve the detection probability. Cooperative sensing consist of a control channel that is used for communicating spectrum sensing results and channel allocation information among cognitive users. The main challenge of cooperative sensing is the development of efficient spectrum information sharing algorithms and the complexity of the approach. The energy detection is the currently approved detection scheme for use in cooperative spectrum sensing because of its simplicity couple with the fact that it does not require a prior knowledge of the PU signals. The energy detector is also the detection method used in this research for the same reasons.

There are three main steps involved in cooperative sensing, these are:

- The fusion centre initiate the cooperation by selecting a channel or frequency band and inform all the cooperating CR users via the control channel to perform local

spectrum sensing on the selected frequency band.

- All the cooperating CRs will sensed the frequency band and report their sensing results to the fusion centre through the control channel.
- Finally, the fusion centre will fused or combined the sensing results of the CRs using a set of predefined rules and come out with a binary decision about the statute of the channel as either busy or idle (vacant). The decision statute is then communicated to the CRs.

There are three major topologies proposed for achieving cooperative sensing in CR network according to their level of cooperation. These are: Centralized coordinated, decentralized coordinated and decentralized uncoordinated techniques.

Centralized coordinated techniques

In centralised sensing, all the secondary users communicate their sensing results on the primary users' activity status to a decision making CR controller commonly called fusion centre (FC) which can be a wired mobile device, access point, base station or another cognitive radio. The fusion centre analyses the information received from the CRs and finally determines the occupancy status of the primary users. The fusion centre does not only disseminate the primary users status to the CRs, it is also responsible for efficient sharing of the available white spaces (spectrum holes) among the CRs. Centralized coordinated spectrum sensing techniques can further be grouped in to two main classes i.e. partially and totally cooperative schemes. In totally cooperative schemes, CRs cooperatively sense the channels and also relay each other information in a cooperative way. In partially cooperative schemes, cooperation is only limited to sensing the channels while the CRs detect the channels independently and communicate their results and decision to the fusion centre.

Fusion rules for centralized coordinated techniques

For the fusion centre to decide whether a channel is busy or idle based on the information received from the CRs about the channel statute, a set of predefined rules are used to draw a conclusion. Some of the fusion rules commonly used are: Hard decision fusion, soft data fusion and quantization data fusion. A brief description of these fusion rules is presented in the next three subsections, Arora and Mahajan (2014).

Hard decision fusion

In the hard decision fusion scheme, each node (CR) sensed the particular channel of interest to determine the present or absent of PU signal. All the cooperating CRs within the cluster will relay their local sensing decision results to the fusion centre via the control channel using one bit which can either be 1 or 0 signifying the present or

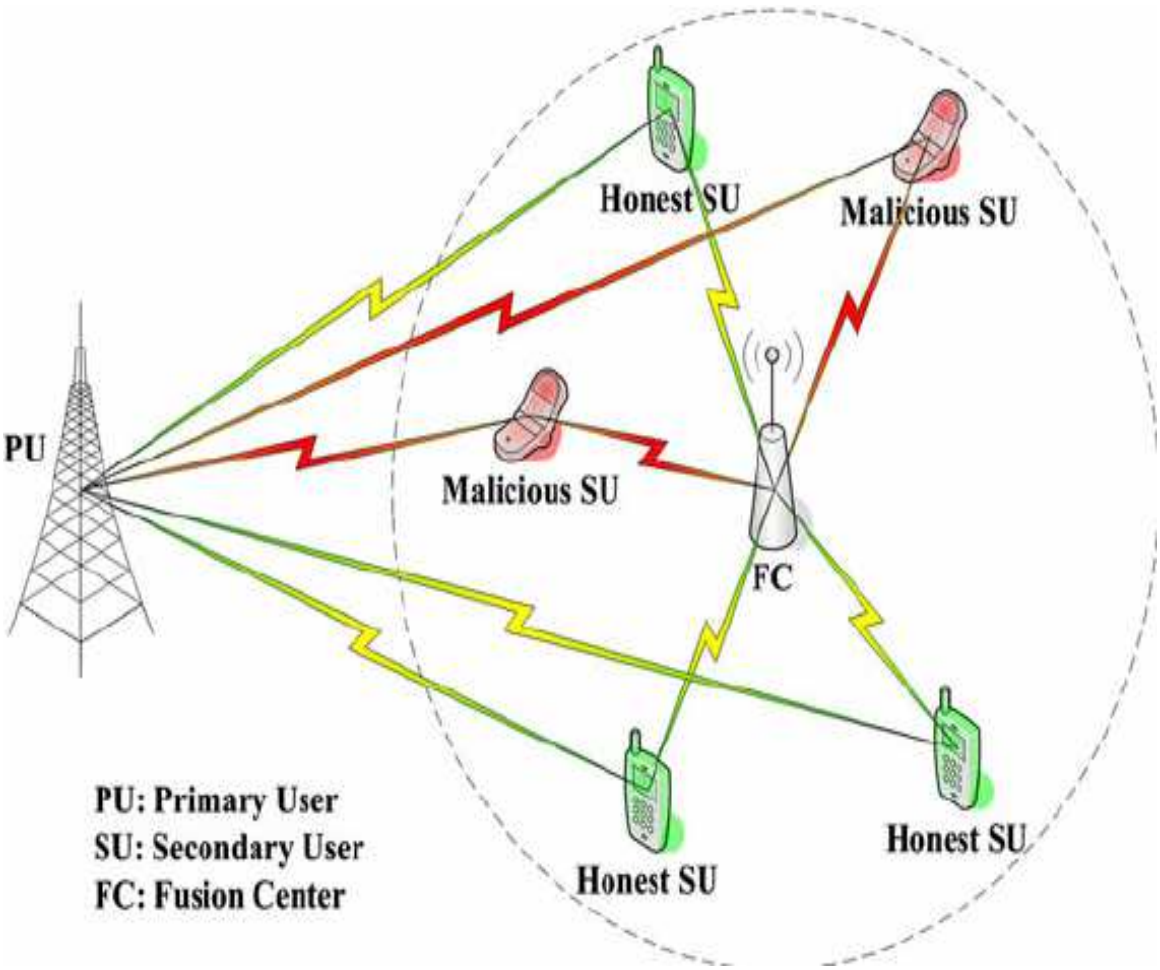

Figure 2.6: Centralized coordinated techniques

absent of PU signal respectively. The fusion centre collect and collate all the binary decision results of the CRs and come out with the final decision about the channel statute using three empirical rules. These rules are the OR, AND and the voting rule. One of the major advantages of hard decision fusion is the reduction in the bandwidth of the control channel since each CRs send its decision using only one bit.

The OR rule

In this rule, if at least one of the cooperating nodes or secondary user local decision is logic 1, then the final decision on the channel statute by the fusion centre will be logic 1. This implies that the channel is currently busy and should not be exploited by cognitive radios. This rule offers high protection to PUs but can result in low spectrum utilization or lost of communication opportunity for CRs in the event of false alarm from any of the cooperating nodes. Any faulty CR among the cooperating nodes whose fault favour false alarm will reduce the overall spectrum efficiency. The OR rule can be

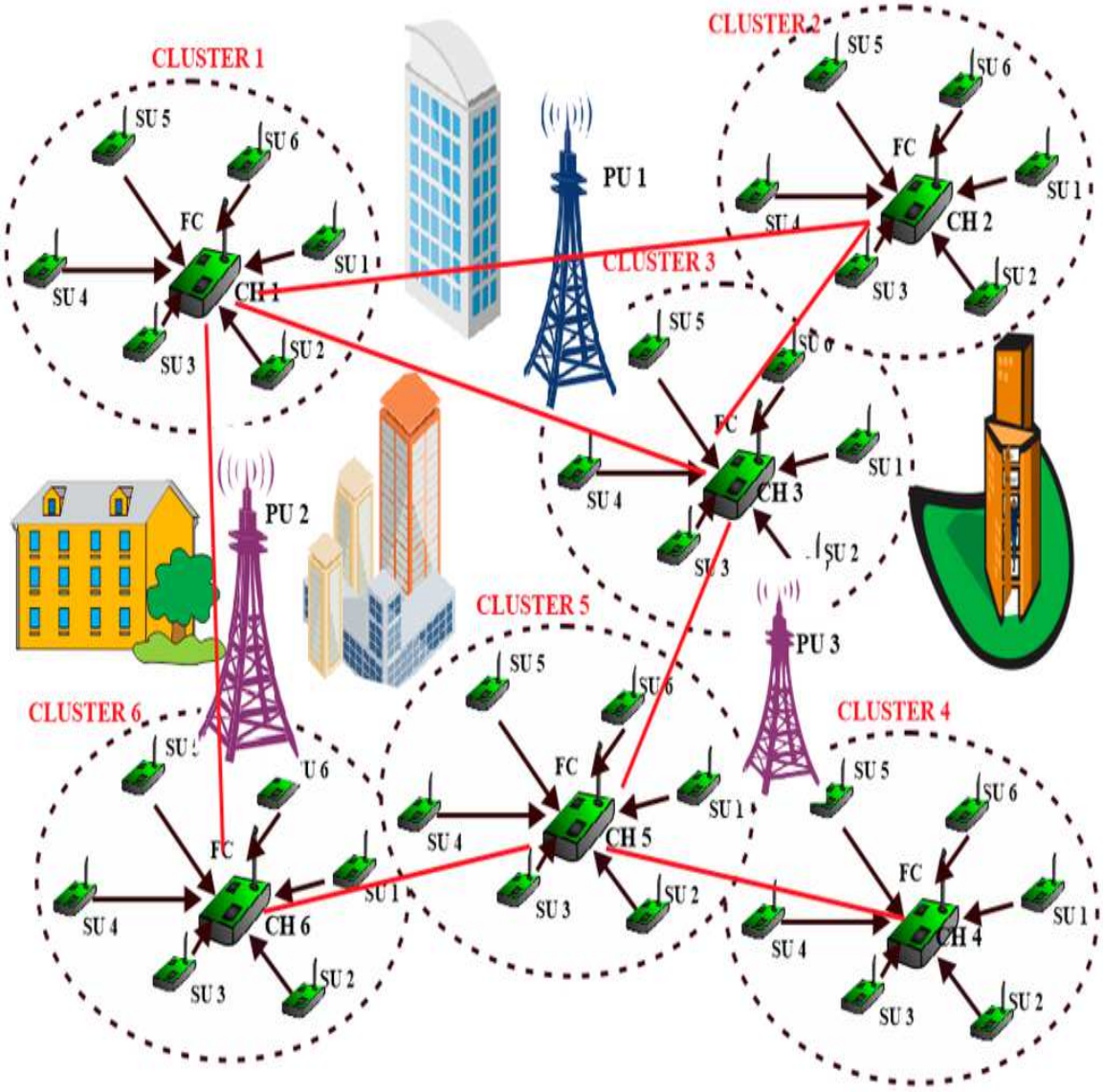

Figure 2.7: Centralized coordinated techniques with cluster heads

model as:

$$\begin{aligned} H_1: &\quad \sum_{i=1}^{M} \vartheta_i \geqslant 1 \\ H_0: &\quad \text{otherwise} \end{aligned} \tag{2.7}$$

Where H_1 represent present of primary users signal while H_0 signify absent of primary user. ϑ_i is the local sensing binary decision of the i_{th} cooperating node (CR), where $i = 1, \cdots, M$ and M is the total number of cooperating nodes. ϑ_i is either 0 or 1. With reference to the OR rule, the CI predictive models developed in this research will minimise the time and energy that will be spent in sensing all the channels before taking a decision. If the channel predicted to be busy by any of the CRs is true, then there is no need of further sensing of the same channel.

The local sensing probability of detection $P_{d,i}$ and false alarm $P_{f,i}$ of i CR is as shown in Equations (2.8) and (2.9) respectively.

$$P_{d,i} = Pr(\vartheta_i = 1|H_1) \tag{2.8}$$

$$P_{f,i} = Pr(\vartheta_i = 1|H_0) \tag{2.9}$$

The cooperating probability of detection $P_{(d,OR)}$ and false alarm $P_{(f,OR)}$ for OR rule is given by Equations (2.10) and (2.11) respectively.

$$P_{(d,OR)} = 1 - \prod_{i=1}^{M}(1 - P_{d,i}) \tag{2.10}$$

$$P_{(f,OR)} = 1 - \prod_{i=1}^{M}(1 - P_{f,i}) \tag{2.11}$$

The AND rule

The AND rule states that the primary user signal is present if all the cooperating nodes detect the presence of signal i.e. if all the local sensing decision of the nodes is logic 1. This rule has the tendency of improving spectrum usage than the OR rule but the primary users stand the risk of suffering from harmful interference in the event of miss detection by any of the cooperating nodes. The AND rule can be formulated as:

$$\begin{aligned} H_1 : &\quad \sum_{i=1}^{M} \vartheta_i \geqslant M \\ H_0 : &\quad \text{otherwise} \end{aligned} \tag{2.12}$$

Where H_1 represent present of primary users signal while H_0 signify absent of primary user. ϑ_i is the local sensing binary decision of the i_{th} cooperating node (CR), where $i = 1, \cdots, M$ and M is the total number of cooperating nodes. ϑ_i is either 0 or 1. The cooperating probability of detection $P_{(d,AND)}$ and false alarm $P_{(f,AND)}$ for AND rule is given by Equations (2.13) and (2.14) respectively.

$$P_{(d,AND)} = \prod_{i=1}^{M} P_{d,i} \tag{2.13}$$

$$P_{(f,AND)} = \prod_{i=1}^{M} P_{f,i} \tag{2.14}$$

The voting rule

This rule states that the primary user signal is present if at least K nodes out of the M cooperating nodes detect a signal during local sensing, where K is the voting threshold in the range $1 \leqslant K \leqslant M$, Horgan and Murphy (2010). The test statistics of the voting rule can be formulated as:

$$\begin{aligned} H_1: &\quad \sum_{i=1}^{M} \vartheta_i \geqslant K \\ H_0: &\quad \text{otherwise} \end{aligned} \tag{2.15}$$

The cooperating probability of detection Q_d and false alarm Q_f for voting rule are as shown in Equations (2.16) and (2.17) respectively.

$$Q_d = \sum_{i=K}^{M} \binom{M}{i} \cdot (1 - P_d)^{M-K} {P_d}^K \tag{2.16}$$

$$Q_f = \sum_{i=K}^{M} \binom{M}{i} \cdot (1 - P_f)^{M-K} {P_f}^K \tag{2.17}$$

where:

$$\binom{M}{i} = \frac{M!}{(M-i)!i!} \tag{2.18}$$

The OR and the AND rules are special cases of the voting rule when $K = 1$ and $K = M$ respectively. The majority rule which state that primary user signal is present if at least halve of the cooperating nodes detect primary user signal, is also another special case of the voting rule when $K \geqslant \frac{M}{2}$.

Soft decision fusion

In soft decision data fusion schemes, all cooperating nodes within a given cluster, perform local sensing on the target channel using energy detection method; and send the energy to the fusion centre without performing any local logical decision regarding the present or absent of primary user signal. The fusion centre or cluster head combined the energies collected from the nodes (CRs) using predefined rules and come up with a binary logical decision regarding the primary user channel statute i.e. either busy (occupy) or idle (vacant). The energy is computed using Equation (2.19). Some of the proposed combining rules in soft data fusion includes: Maximal ratio combination (MRC) rule, square law combination rule and selection combination rule. In comparison with the hard fusion scheme, the soft data fusion is more reliable and accurate than the hard fusion scheme. One of the major drawback of the soft fusion scheme is the fact that it required a control channel with a large bandwidth to transmit all the energies

from the nodes to the fusion centre.

$$E_i = \sum_{n=1}^{N} y(n)^2 \tag{2.19}$$

Where E_i is the energy of the i^{th} CR, N is the number of samples while $y(n)$ is the received sample signal.

Square law combination rule: In this rule, all the energies E_i of the cooperating nodes sent to the fusion centre are summed up and compared with a predefined threshold for decision making. Based on this rule, primary user signal is present if the sum of the energies E is greater than the threshold T otherwise the channel is idle.

$$E = \sum_{i=1}^{M} E_i \tag{2.20}$$

The test statistics of square law combination is given as:

$$\begin{aligned} H_1 : &\quad E > T \\ H_0 : &\quad \text{otherwise} \end{aligned} \tag{2.21}$$

Where H_1 and H_0 are the two hypothesis for busy and idle respectively.

Maximal combination rule: In the maximal combination rule, the energies of the cooperating CRs sent to the fusion centre are weighted with a normalized weight. The sum of the weighted energies is compared with the threshold for decision making on the primary user channel statute. The weights are chosen in proportion to the signal to noise ratio (SNR) of each CR as reported to the fusion centre.

$$E = \sum_{i=1}^{M} w_i E_i \tag{2.22}$$

Where w_i and E_i are the weight and energy of node i respectively, while E is the sum of the weighted energies of the CRs. The sum of the weights W is constrained to 1 i.e.

$$W = \sum_{i=1}^{M} w_i = 1 \tag{2.23}$$

The test statistics is given as:

$$\begin{aligned} H_1 : &\quad E > T \\ H_0 : &\quad \text{otherwise} \end{aligned} \tag{2.24}$$

Selection combination: The energy of the CR with the highest SNR among the cooperating nodes is selected. The decision on the statutes of the PU channel (busy or idle) is determined by the energy of the selected CR and the decision threshold. If noise is assumed to be AWGN, then the probability of detection Q_d and false alarm Q_f for

selection combination scheme can be expressed in terms of the signal to noise ratio υ of the selected CR and the decision threshold T as shown in Equations (2.26) and (2.27) respectively.

$$\upsilon = \max(\upsilon_1, \upsilon_2 \cdots \upsilon_M) \tag{2.25}$$

$$Q_d = Q_\varphi(\sqrt{2\upsilon}, \sqrt{T}) \tag{2.26}$$

$$Q_f = \frac{\Gamma(\varphi, T/2)}{\Gamma(\varphi)} \tag{2.27}$$

Where φ is the product of the bandwidth and time, $\Gamma(.)$ and $\Gamma(.,.)$ is incomplete and complete gamma functions respectively. Q_φ is the generalized Marcum Q function.

Decentralized coordinated technique

In this case, nodes build a network without a fusion or control centre. Clustering or gossiping algorithms is one of the algorithms suggested to be adopted by decentralized coordinated spectrum sensing. In this proposed algorithm, nodes (CRs) auto coordinate themselves and form a cluster. Coordination among the nodes is improved via the use of a control channel which is implemented by underlay ultra wide band (UWB) channel or dedicated frequency channel, Subhedar and Birajdar (2011a), Sharma (2013).

Decentralized Uncoordinated techniques

There is no one form of coordination or cooperation that exists among the cognitive users (CRs). Each CRs independently sense channels in search of spectrum holes and used any spectrum hole found without informing any CR. In the same way if any CR detect the presence of PU, it will simply vacate the channel without relating the information to other CRs. Decentralized uncoordinated technique is very weak and inefficient in maximizing the usage of the limited spectrum resources. Is also prompt to causing interference to PU as a result of miss detection.

2.2.5 Interference Based Detection

Interference based detection (IBD) provides the means by which the cognitive radios can coexist with the primary users in an underlay cognitive radio scheme. There are two approaches toward the realization of IBD, these are: Interference temperature measurement and primary receiver detection. The CI RF power prediction models developed found their widest application in interference based detection schemes. The a priori knowledge of the RF power of the channels will make the CRs to dynamically increase or decrease their transmitted power thereby increasing the spectrum utilization efficiency without causing harmful interference to the PUs.

Interference temperature measurement

In this schemes, CRs are allow to coexist with the PUs within a given frequency bands and at a given geographical location; by controlling their transmitted power in order to avoid any harmful interference to the PUs. The CRs are allowed to transmit along with the PUs as long as the aggregate interference at the PU receivers does not exceed a given threshold called the interference temperature limit. The interference temperature refer to the temperature equivalent of the aggregate RF power at the PUs receiver antenna per unit bandwidth as shown in Equation (2.28). This aggregate power is the sum total of RF power at the receiving antenna of the PU resulting from both PU and CR transmitters, and noise. The interference ceiling (threshold) for each band is set by USA Federal Communication Commission (FCC), Communication (2003).

$$T_I(f_c, B) = \frac{P_I(f_c, B)}{KB} \tag{2.28}$$

where $P_I(f_c, B)$ is the average interference power in Watts centred at frequency f_c, B is the bandwidth in Hz, K is Boltzmann constant in Ws/K while $T_I(f_c, B)$ is the interference temperature in Kelvin.

Primary receiver detection

The presence of PUs is meant to be detected by sensing the leakage power of the PU receiver local oscillator. This proposed spectrum sensing method is currently impractical due to two major reasons: It will be very hard to detect directly the local oscillator leakage power as it varies with time and age of the receiver, secondly, since the leakage power is very low, direct detection by the CRs over significant distant will require a CR with very high sensitivity which may be difficult to attain.

The next section presents a brief review of some areas where CI methods were applied in cognitive radio network and the comparison with the current work.

2.3 Review of some applications of CI in CR

There are different types and variants of Computational Intelligence (CI) and machine learning algorithms that can be used in CR such as genetic algorithms for optimization of transmission parameters Rondeau et al. (2004), swarm intelligence for optimization of radio resource allocation Udgata et al. (2010), fuzzy logic system (FLS) for decision making, Matinmikko et al. (2013), Giupponi and Perez (2008), Jethro et al. (2010), neural network and hidden Markov model for prediction of spectrum holes; game theory, linear regression and linear predictors for spectrum occupancy prediction Chen and Oh (2013), Bayesian inference based predictors, etc. Some of the CI methods are used for learning and prediction, some for optimization of certain transmission parameters while others for decision making, Azmi (2011). TV idle channels prediction using ANN

was proposed by Winston et al. (2013), however, data were collected only for two hours everyday day (5pm to 7pm) within a period of four weeks, this is not sufficient to capture all the various trends associated with TV broadcast. Also, identifying the idle channels does not depict any spatial or temporal information of the expected noise and/ or level of interference based on the channels history which is vital in selecting the channels to be used among the idle channels. Spectrum hole prediction using Elman recurrent artificial neural network (ERANN) was proposed, Taj and Akil (2011). It uses the cyclostationary features of modulated signals to determine the presence or absence of primary signals while the input of the ERANN consists of time instances. The inputs and the target output used in the training of the ERANN and prediction were modelled using ideal multivariate time series equations, which are often different from real life RF traffics where PU signals can be embedded in noise and/ or interfering signals. Traffic pattern prediction using seasonal autoregressive integrated moving-average (SARIMA) was proposed for reduction of CRs hopping rate and interference effects on PU while maintaining a fare blocking rate, Li and Zekavat (2008). The model (SARIMA) does not depict any information about the expected noise power.

Fuzzy logic (FL) is a CI method that can capture and represent uncertainty. As a result it has been used in CR research for decision making processes. In Matinmikko et al. (2013) an FL based decision-making system with a learning mechanism was developed for selection of optimum spectrum sensing techniques for a given band. Among these techniques are matched filtering, correlation detection, features detection, energy detection, and cooperative sensing. Adaptive neural fuzzy inference system (ANFIS) was used for prediction of transmission rate, Hiremath and Patra (2010). This model was designed to predict the data rate (6, 12, 24, 36, 48 and 54 Mbps) that can be achieved in wireless local area network (WLAN) using a 802.11a/g configuration as a function of time. The training data set was obtained by generating a random data rate with an assigned probability of occurrence at a given time instance, thus forming a time series. In this study, real world RF data was not used. More importantly, the research did not take into account the dynamic nature of noise or interference level which can affect the predicted data rates. Semi Markov model (SMM) and continuous-time Markov chain (CTMC) models have also been used for the prediction of packet transmission rates, Geirhofer et al. (2009). This avoids packet collisions through spectrum sensing and prediction of temporal WLAN activities combined with hoping to a temporary idle channel. However, SMM are not memory efficient, neither was there any reference made to the expected noise level among the inactive (idle) channels to be selected. An FL based decision system was modelled for spectrum hand-off decision-making in a context characterized by uncertain and heterogeneous information Giupponi and Perez (2008) and fuzzy logic transmit power control for cognitive radio. The proposed system was used for the minimization of interference to PU's while ensuring the transmission rate and quality of service requirements of secondary users, Tabakovic et al. (2009). The researcher did not, however, include any learning from past experience or historical data. An exponential moving average (EMA) spectrum sensing using energy prediction was implemented by Lin et al. (2009). The EMA achieved a prediction average mean

square error (MSE) of 0.2436 with the assumption that the channel utilization follow exponential distribution with rate parameter $\lambda = 0.2$ and signal to noise (SNR) of 10dB; RF real world data was not used in their study.

This research demonstrates the use of support vector regression (SVR) and ANN trained using the proposed hybrid nested algorithms for prediction of real world RF power of some selected channels within the GSM band, VHF and UHF bands. An optimised ANN is designed by combining the global search advantages of population based metaheuristics optimization algorithms and the local search advantages of single solution back propagation algorithm to train the network. The SVR is trained using an ad-hoc nested algorithm whose outer layer is a metaheuristic, while the inner layer is a convex optimization algorithm. A combined prediction model consisting of ANN and SVR models was found to yield a more promising result with an average improvement of 46%. The initial weights of the ANN were evolved using metaheuristic algorithm after which the ANN was trained (fine tuned) more accurately using back-propagation algorithms. This methodology demonstrates the application of previously acquired real world data to enhance the prediction of RF power to assist the implementation of CR applications.

2.4 Metaheuristics Optimization Algorithms

In this section, a general review of Computational Intelligence (CI) techniques are presented, with emphasis on the ones used in this study. Despite the gross advancement in CI techniques, the definition of CI is still ambiguous as there is no generally acceptable definition of CI since the term was first used in 1990 by the IEEE neural networks council, Siddique and Adeli (2013). Artificial Intelligence (AI) methods are systems that simulate human intelligent behaviour which need exact and complete knowledge representation of the problem, model, process or system. Unfortunately, many real world systems cannot be modelled with absolute precision, and complete knowledge which are the two major requirements of AI. This is due to the fact that real world systems are characterized with uncertainties, vagueness, imprecision, nonlinearity and unanticipated changes in the environments or process, and the model objectives. These uncertainties make some aspects of real world systems to be difficult to model due to lack of exact and complete knowledge representation. Hence there is a need for advanced decision making systems that can achieved certain acceptable performance in the mist of uncertainties, incomplete knowledge and severe changes in system operating conditions. CI systems are intelligent computing systems with human-like capabilities to make decision and future prediction based on an information with uncertainties and imprecision. They have the ability to learn and adapt i.e. to dynamically change or evolve their parameters or structure in order to best meet their objective target or goal, Fogel (1995). Both AI and CI are two broad views of machine intelligence as shown in Figure 2.8 with their associated alliance, Siddique and Adeli (2013). This structure shows hard computing (HC) schemes as a subset of AI while soft computing (SC) tech-

nique as a large subset of CI. Among other techniques that belong to SC framework are Fuzzy Systems (FS), Neural Networks (NN) and Evolutionary Algorithms (EA). Imprecision and uncertainties are undesirable features in hard computing while in soft computing they are they the most prioritized features, Zadeh (1998). Soft computing are fundamental methodologies that form the bedrock for designing intelligent systems. In a nutshell, CI is a combination of soft computing and numeric processing.

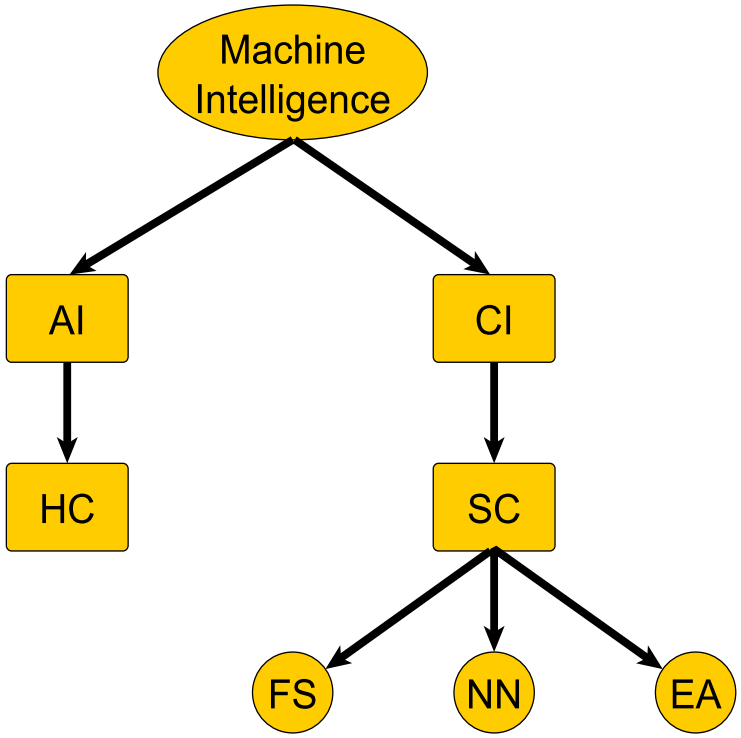

Figure 2.8: Machine intelligence techniques with their alliance

CI is an interdisciplinary field with each discipline having a unique approach to CI in addressing problems within their scope. These broad interdisciplinary techniques include: Fuzzy logic, neural network, learning theory, swarm intelligence, probabilistic methods (models) and evolutionary computing as shown in Figure 2.9. Four of these techniques (i.e. neural networks, learning theory, swarm intelligence and evolutionary computing) are used in this research to develop optimized systems that predict spectrum holes and RF power of primary users (PUs) channels for effective spectrum utilization. These four are detailed in this thesis while others are briefly explained where applicable. Significant improved performance of the prediction models were achieved using hybrid models consisting of a combination of multiple CI techniques.

2.4.1 Evolutionary Algorithms

Evolutionary Algorithms (EAs) has played a vital role in solving modal and complex multimodal nonlinear optimization problems since 1960s, drawing their roots from the concept of evolution introduced by Sewell Wright in 1932, Wright (1932). This broad

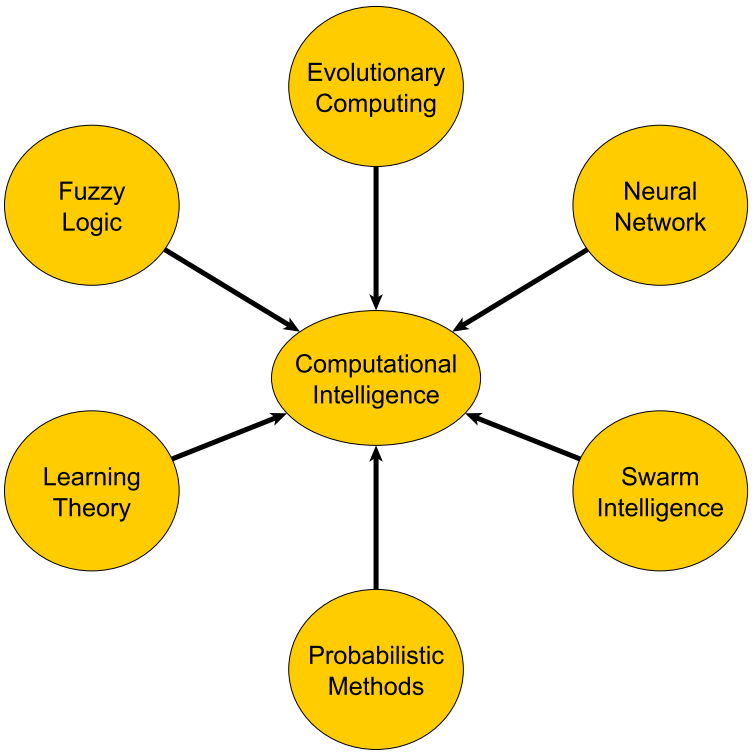

Figure 2.9: Computational intelligence techniques

field of intelligent optimization frameworks inspired by natural phenomena, has grown over the years in scope, diversity and areas of applications with increasing demand for more. EAs are metaheuristic optimization algorithms with many variants which include: genetic algorithms (GA), evolutionary strategy (ES), evolutionary programming (EP) and genetic programming (GP), Eiben and Smith (2003), Jong (2006). EAs has been proved to be an effective tool for solving multi-objective optimization problems, Matthews et al.; 2011). These are population based randomization solution search space algorithms. Though the search procedure or methodology adopted by each of these algorithms for solving a given problem is different, the underlying principle or idea for all the EAs techniques is the same i.e. they only differ in technical details, Eiben and Smith (2003). The general framework of the EAs is as shown in Figure 2.10 and Algorithm 2.1. With reference to Figure 2.10, the general components of EAs can be summarized as follows:

- **Representation:** The first step in EAs is the representation of the candidate solutions in the population. The individuals can be represented as vector of binary strings, real numbers, decision rules, functions, alternative ways of configuring something, etc. The lower and upper bound of each gene (element) in the vector is set in order to form the decision or search space. The candidate solution in this research are represented by a vector of real numbers.

- **Evaluation function:** A means of computing the fitness of each candidate solution should be provided. The evaluating function is referred to as the fitness

function which is the basis for selection and a means of improving the quality of the population. The term "fitness" refers to the assessment quality of an individual (candidate solution) in solving the problem. Is the measure of how close the candidate solution is able to solve the problem as compared with the set objective aim, hence the term is also referred to as objective fitness function. This is the metric used to evaluate the performance of each candidate solution of the metaheuristics used in this research.

- **Population:** The population size and the dynamics involved are specified. In most EAs, the population size is constant through out the evolutionary process. The population size refers to number of possible solutions at any generation or iteration. The number of different possible solutions present, is one way of evaluating the diversity of the population. The choice of population of appropriate size is problem dependent.

- **Parent selection mechanism:** This is similar to survival selection but is only done once within the initialization phase. Parents selection mechanism are often stochastic contrary to survival selection which are deterministic. Fitness bias approaches can also be used for parent selection. Parent selection mechanism are used to select parents that will serve as the initial seed to start the evolutionary process. Proper selection of the initial seeds (parents) over the decision space is key to the success of the evolutionary process. In this research, the initial parent are randomly generated to cover the decision space as much as possible using normal distribution random number generator.

- **Variational operators:** These operators are meant to create new interesting solutions from the current existing solutions (parents). There are two strategies for implementing variational or reproductive operators in EAs, these are mutation and Recombination. Mutation involve the perturbation of an existing solution (one parent) to produce new solution or provisional offspring called the mutant while recombination hybridized or combined two or more parents (existing solutions) to produce new solution. EAs that uses both strategies (mutation and recombination) in their framework are often better than those that uses only one of the two, Jong (2006). Some EAs such as differential evolution use mutation which involves at least three parents, and crossover as its variational operators as detailed in Section 2.6.1. Reproduction variation is the major source of exploration in EAs, hence the level of exploration can be controlled via the choice of reproduction mechanism. These operators are designed to effectively manipulate and extract meaningful building blocks from the existing solutions in order to produce better solutions.

- **Survival selection:** When the offspring are birthed, the survival mechanism is used to determined which individuals (i.e. among the parents and the offspring) should be allowed to proceed to the next generation. This choice is often based on the fitness function or quality of the individuals. The fitness bias selection

scheme is used in overlapping generation models where both parents and offspring are allowed to seed the next generation subject to their fitness. The fitness biased selection includes:

1. **The greedy selection scheme:** In this selection scheme, an offspring is compared with its parent and the fittest is selected.

2. **The ranking selection:** In this case all the parents and the offspring are ranked according to their fitness and the topmost are selected according to the size of the population, Eiben and Smith (2003).

3. **Ranking proportional selection:** This is another form of ranking selection where the pool of the population are rank according to their fitness but instead of selecting the topmost, each potential solution is assigned a probability value in decreasing order starting from the most fittest to the least such that the fitter ones has better chance of been selected. The probability of selecting a solution with rank i is given by Equation (2.29), Jong (2006).

$$prob(i) = (\frac{2}{m} - \in) - (\frac{2}{m} - 2 \in) * \frac{i-1}{m-1} \tag{2.29}$$

Where m is the number of parents, $\frac{2}{m}$ and $\in$ are the probability mass for selecting the most fittest candidate and the worst candidate respectively. The sum of the probability mass $prob$ is 1.

4. **Truncation selection:** This is the most elitist selection scheme where the pool of the candidates solutions are ranked based on their fitness starting from the most fittest candidate to the least and the top k candidates are selected for breeding and to seed the next generation. If r is the pool population, then $k < r$.

5. **Fitness proportion selection:** In this selection approach, the selection pressure is self-adapted with modest programming effort. This mechanism is realised by defining the probability $Prob_{j,G}$ of selecting candidate j with fitness $Fitnss_{j,G}$ at generation G with respect to the total fitness $Fitness_{T,G}$ as shown in Equations (2.30) and (2.31) for maximization and minimization respectively.

$$Prob_{j,G} = \frac{Fitnss_{j,G}}{Fitness_{T,G}} \tag{2.30}$$

$$Prob_{j,G} = 1 - \frac{Fitnss_{j,G}}{Fitness_{T,G}} \tag{2.31}$$

6. **Tournament selection:** While ranking and truncation selection methods required the sorting of the population pool by fitness which my be computationally expensive for large population; the tournament selection can be used to mitigate this cost. In tournament selection, the pool is unsorted and the individuals that win a tournament of size k are the ones that will be selected. A tournament is formed by selecting k individuals from the pool using a uniform probability distribution with replacement, and the candidate with the best fitness is crowned the

winner of the tournament. The only way the best genotype (candidate solution) will not win the tournament is when it is not selected. For a tournament of size k, the proportion to be selected in the next generation P_{G+1} can be expressed in terms of the proportion selected in the current generation P_G as:

$$P_{G+1} = 1 - (1 - P_G)^k \tag{2.32}$$

For binary tournament, i.e. $k = 2$ equation (2.32) become:

$$P_{G+1} = 1 - (1 - P_G)^2 = 2P_G + P_G^2 \tag{2.33}$$

7. **Age biased selection:** In the non-overlapping generation model, only the offspring are allowed to proceed to the next generation, the survival selection mechanism often used in this model is age biased. Here the offspring replaces their parents after every generation.

Selection is the main source of exploitation in EAs, thus the choice of the selection mechanism controlled the degree of exploitation. A good EA should find an efficient way of enhancing exploration of unexplored regions within the decision space and effective exploitation of the regions already explored. The optimum choice of the selection schemes are problem dependent and can not be generalized. The greedy selection scheme is used in this research during the optimization of the RF prediction models, as discussed in Section 2.6.1.

- **Termination:** Since EAs are stochastic in nature, there is no guarantee of achieving the exact optimum solution even for problems with known optimums. Hence there is need to set a realistic stopping criteria to terminate the evolutionary process. Some commonly used termination schemes are:

  1. Elapse of maximum CPU allowed time
  2. Number of iteration or fitness function evaluation reaches a set limit.
  3. No improvement in fitness function within a given threshold for a given period of time or number of iteration.
  4. Drop of population diversity under a given threshold.

  Three terminations criteria were used for the algorithms used in this study, these are: The maximum number of fitness function evaluation or the error is within a given limit or the optimization will stop when there is no improvement in the fitness function for a given number of iteration.

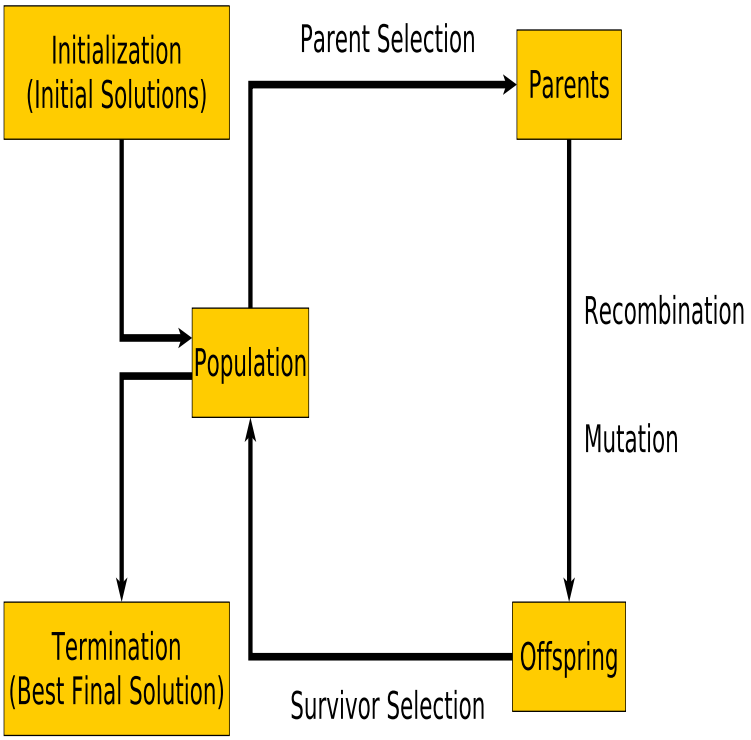

Figure 2.10: General framework of evolutionary algorithms

Algorithm 2.1 Evolutionary algorithms framework psuedocode

INITIALIZATION
Initialise the population uniformly over the decision space with random potential candidate solutions
Evaluate the fitness of each candidate
PARENTS SELECTION
while Termination condition is not reached **do**
  RECOMBINATION: Combine two or more parents
  MUTATION: Mutate the resulting offspring
  EVALUATE: Compute the fitness of new candidates
  SURVIVAL SELECTION: Select individuals for next generation
end while

2.5 Predictive Models

Prediction models based on artificial neural network and support vector machine framework are used in this research for spectrum hole and RF power prediction. The theory of the prediction models, and the approaches adopted are discussed in this section.

2.5.1 Artificial Neural Network

Artificial Neural Networks (ANN) are massive parallel computing systems composed of large number of simple processing elements (processors) with many interconnections operating in parallel. These elements are inspired by the structure and function of

biological nervous systems such as the brain, Haykin (2008), Jianli (2011), Suzuki (2011), Huemer and Elizondo (2010), Huemer et al. (2009; 2008b;a), Gongora and Elizondo (2005). The information processing units are called neurons which are interconnected together via a synaptic weights, and working together in parallel and unison to solve specific problems. The mathematical model for ANN was first initiated by Warren McCulloch and Walter Pitts in 1943, Warren and Walter (1943), this model form the basis upon which the conventional ANN models were built. This model was mainly designed for binary classification, the mathematical representation of the firing rule is as shown in Equation (2.34) and depicted graphical in Figure 2.11

$$F = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} w_i x_i \geqslant T \\ 0 & \text{otherwise} \end{cases} \tag{2.34}$$

Where T is a predefined threshold, x is multi-variant input vector while w is the corresponding weight vector. In general, the term weight in ANN referred to the strength of connection between two neurons which is the measure of the weight of information flow from the output of one of the connected neurons to the input of the another. $i = 1 \cdots n$ where n is the number of input variables.

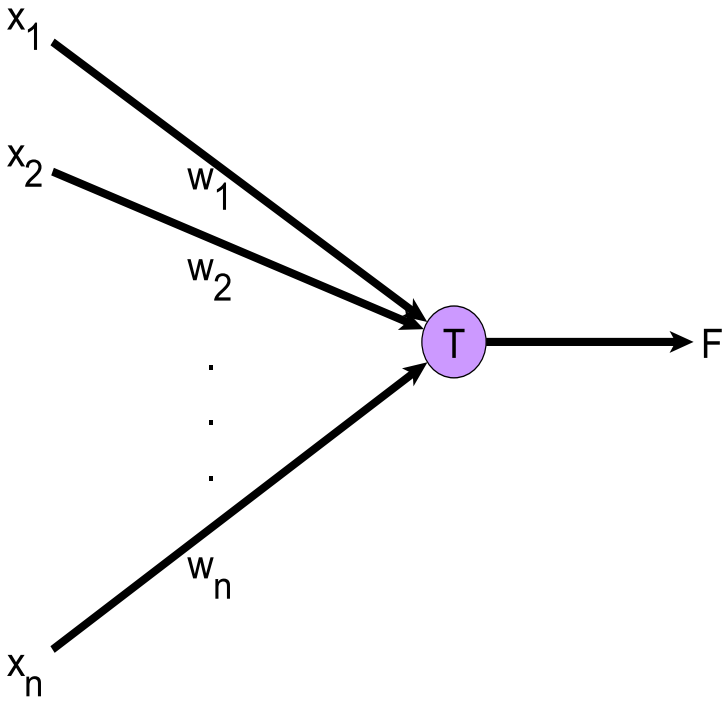

Figure 2.11: McCulloch and Pitt ANN

The McCulloch and Pitt's neuron model was improved by Frank Rosenblatt in 1958 who introduce another constant parameter to the neuron model called the bias, Frank (1958). The ANN model invented by Rosenblatt is called the perceptrons which

is also used for binary classification, the model is as shown in Figure 2.12, represented by Equation (2.35). This model is found to depict a good result when used to classify linearly separable continuous set of multi-variants inputs into any of the two classes. The application of the perceptron is unfortunately limited and often used for solving linear separable classification problems.

$$F = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} w_i x_i - b \geqslant T \\ 0 & \text{otherwise} \end{cases} \tag{2.35}$$

where the constant parameter b is the bias.

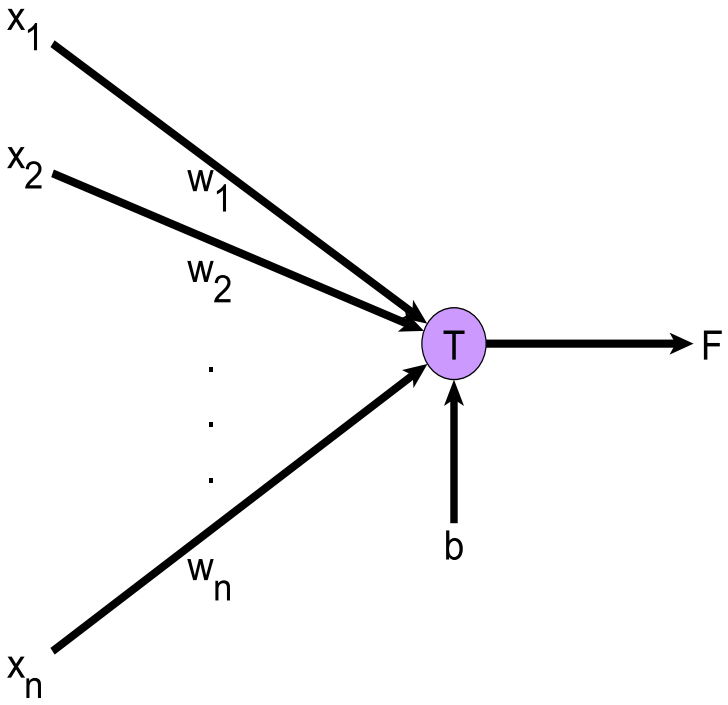

Figure 2.12: The perceptron

The output of the neuron model shown in Figure 2.12 is always driven to either of the two possible states or limits (i.e. 0 or 1). In order to allow for varying output conditions, a non-linear activation function $f(.)$ is often included in the neuron model as shown in Figure 2.13, where z is called the activation given by Equation (2.36) and y is the output of the neuron, Equation (2.37).

$$z = \sum_{i=1}^{n} w_i x_i + b \tag{2.36}$$

$$y = f(z) \tag{2.37}$$

Activation function The activation functions play a vital role in ANN, unfortunately optimum choice of activation functions for each neuron in a given network are

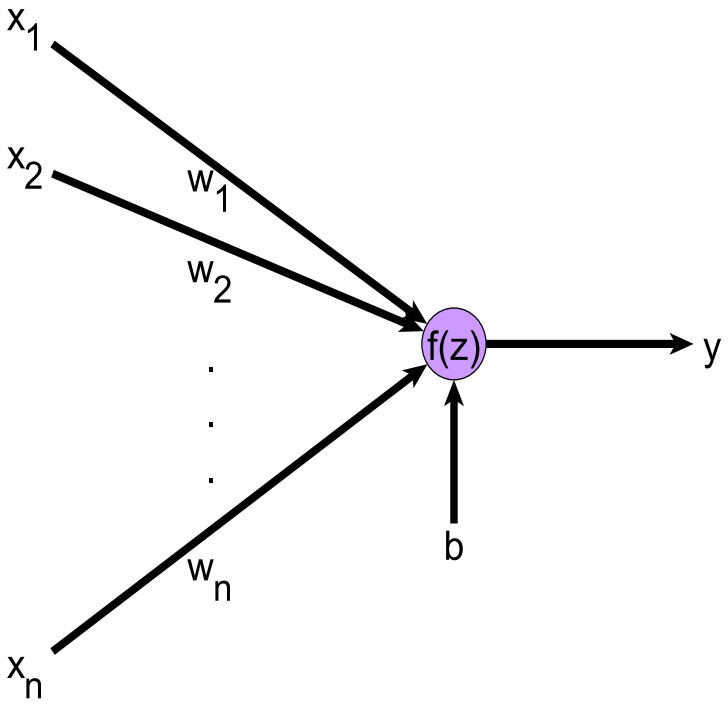

Figure 2.13: The perceptron with activation function

problem dependent and can not be generalised. Some of the commonly used activation functions are: linear function, truncated linear function, ramp function, tansigmoidal function, logistic function and Gaussian radial basis function. Note, for all the activation function shown in Equations (2.38) to (2.45), z is as shown in Equation (2.36).

For this study, due to the dynamic nonlinearity often associated with RF traffic patterns, coupled with random interfering signals or noise resulting from both artificial and natural sources, a fully connected multilayer perceptron (MLP) ANN with two hidden layers was used in this study. The input layer was cast into a high dimensional first hidden layer for proper features selection. The activation functions used in the two hidden layers are nonlinear tansigmoidal (hyperbolic tangent) functions, Equation (2.43), and a linear symmetric straight line, Equation (2.38) is used for the output activation function. For the probabilistic ANN used for spectrum hole prediction presented in Section 4.5.2, the output activation function is logistic sigmoidal function, Equation (2.44). Implementation with other activation functions were also adopted, but this choice gave a better promising results. The nonlinear hyperbolic tangent functions introduced a nonlinear transformation into the network. The hidden layers serve as a feature detector i.e. during the training; they learn the salient attributes (features) that characterizes the training data.

- Linear function:

$$y = f(z) = K \cdot z \tag{2.38}$$

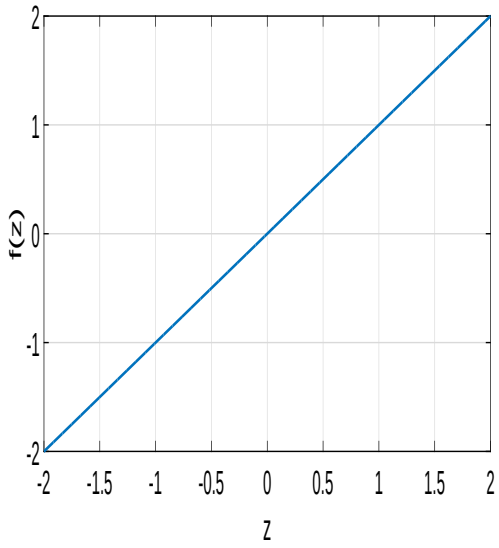

Figure 2.14: Linear function

- Truncated linear function:

$$y = f(z) = \begin{cases} K \cdot z & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.39}$$

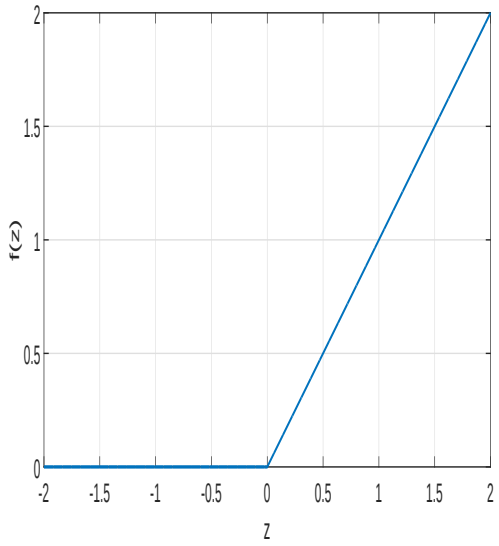

Figure 2.15: Truncated linear function

- Ramp function:

$$y = f(z) = \begin{cases} max\ y & \text{if } z > \text{ upper limit} \\ K \cdot z & \text{if upper limit } \geqslant z \geqslant \text{ lower limit} \\ min\ y & \text{if } z < \text{ upper limit} \end{cases} \tag{2.40}$$

The positive constant K in Equations (2.38) to (2.40) is the amplification factor or gradient of the linear function.

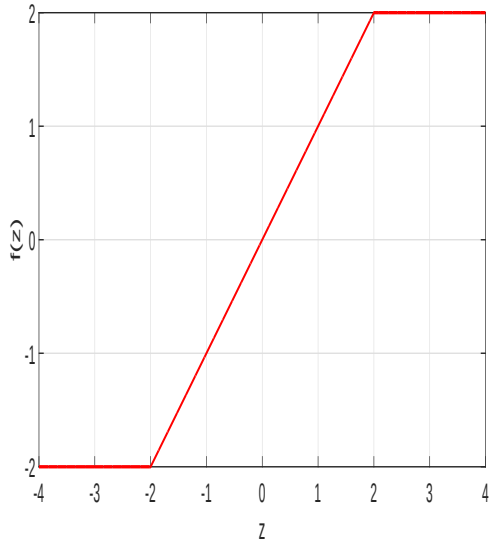

Figure 2.16: Ramp function

- Step function:

$$y = f(z) = \begin{cases} +1 & \text{if } z \geqslant 0 \\ -1 & \text{otherwise} \end{cases} \tag{2.41}$$

$$y = f(z) = \begin{cases} 1 & \text{if } z \geqslant 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.42}$$

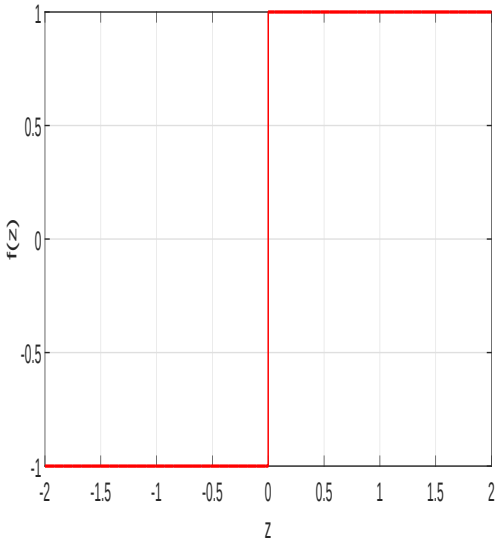

Figure 2.17: Step function with +1 and -1 limits

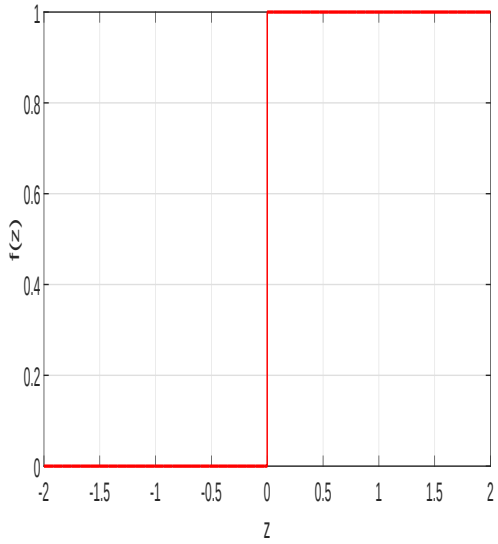

Figure 2.18: Step function with 1 and 0 limits

- Hyperbolic tangent function (Tansigmoidal function):

$$y = f(z) = \frac{1 - e^{-z}}{1 + e^{-z}} \tag{2.43}$$

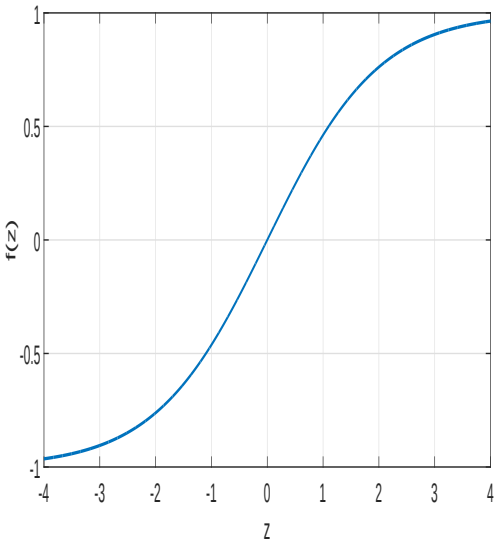

Figure 2.19: Hyperbolic tangent function

- Logistic sigmoidal function:

$$y = f(z) = \frac{1}{1 + e^{-z}} \tag{2.44}$$

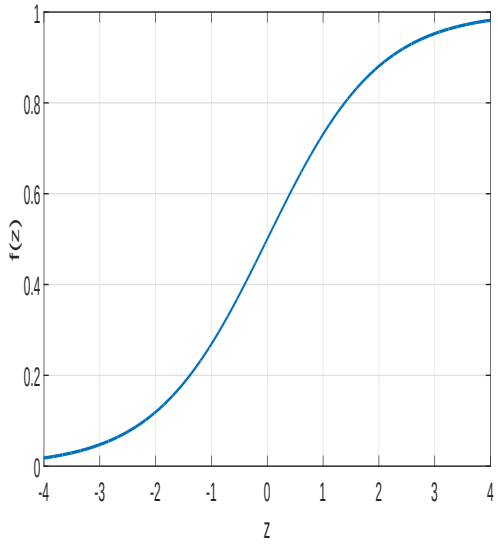

Figure 2.20: logistic function

- Radial basis function:

$$y = f(z) = e^{-z^2} \tag{2.45}$$

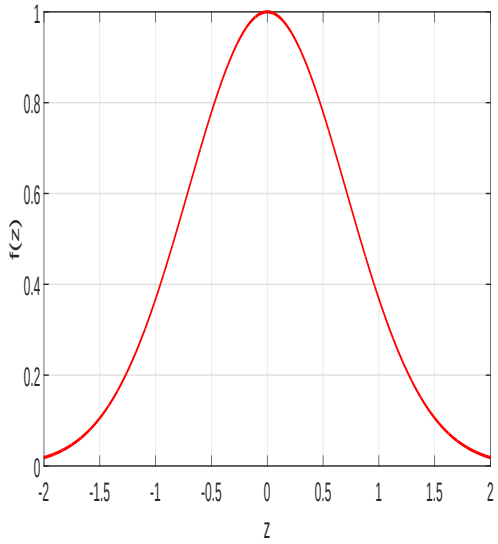

Figure 2.21: Radial basis function

Artificial neural network architecture

Artificial neural network architectures can be broadly classified as either feed forward or recurrent type, Haykin (2008). Each of these two classes can be structured in different configurations. A feed forward network is one in which the output of one layer is

connected to input of the next layer via a synaptic weight, while the recurrent type may have at least one feedback connection or connections between neurons within the same layer or other layers depending on the topology (architecture). The training time of the feed forward is less compared to that of the recurrent type but the recurrent type has better memory capability for recalling past events. Four ANN topologies were considered in this study, i.e. feed forward (FF), cascaded feed forward (CFF), feed forward with output feedback (FFB), and layered recurrent (LR) ANN. Two or more neurons can be combined in one layer, and one or more layers can be combined to form the ANN model, Demuth and Beale (2002). The diagram of single layer and two layers feed forward multilayer perceptron ANN are as shown in Figure 2.22 and 2.23 respectively. Note, the input layer does not contain processing elements rather it is the connection between the ANN and the input data.

All the ANN topologies used in this studies consist of two hidden layers, and parallel outputs or one output as shown in Figure 2.23. The inputs of the ANN consist of recent past RF power samples, frequency or channel, and time domain data of varying rates of change i.e. second, minute, hour, week day (1 to 7), date day (1 to at most 31), week in a month (1 to 5), and month (1-12) while the output gives the power in Decibels (dB). Each input of the time domain, enables the ANN to keep track with the trend of RF power variation as a function of that particular input. The current RF power is thus modelled as a nonlinear function of recent past RF power samples, frequency, and current time. The number of past RF power samples to be used as part of the input of the ANN for reliable prediction and efficient computer resource management was obtained experimentally using sensitivity analysis test as detailed in Section 4.3. The proposed prediction models has proved to be effective for CR application especially where the CR has no prior knowledge of the primary users' signal type, modulation, signal to noise ratio etc; as none of these RF power related parameters are used as part of the input of the ANN. More so primary users may not be willing to release their signals information to cognitive users.

ANN learning algorithms

The procedure or means used in updating the weights and bias of ANN in order to minimise the empirical risk function is called learning. There are three learning procedures used in ANN, these are:

1. Supervised learning

In supervised training, the training data set used in training the ANN consist of input / output data. The inputs are applied to the ANN and the output of the ANN are compared with the target output. The difference between the ANN output and the target output gives the error. The error is used to adjust the weights and biases of the ANN to move the ANN outputs closer to the targets, thus minimising the error. This is also referred to as learning with a teacher where the teacher provide the network with the desired target output vector for each input vector. The training is repeated

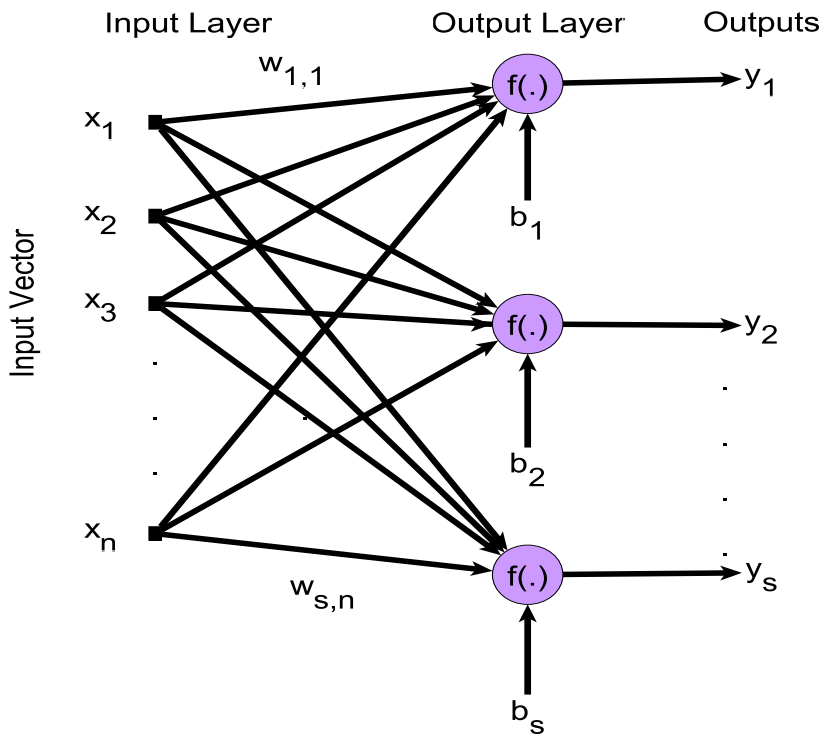

Figure 2.22: Single layer multiple outputs ANN

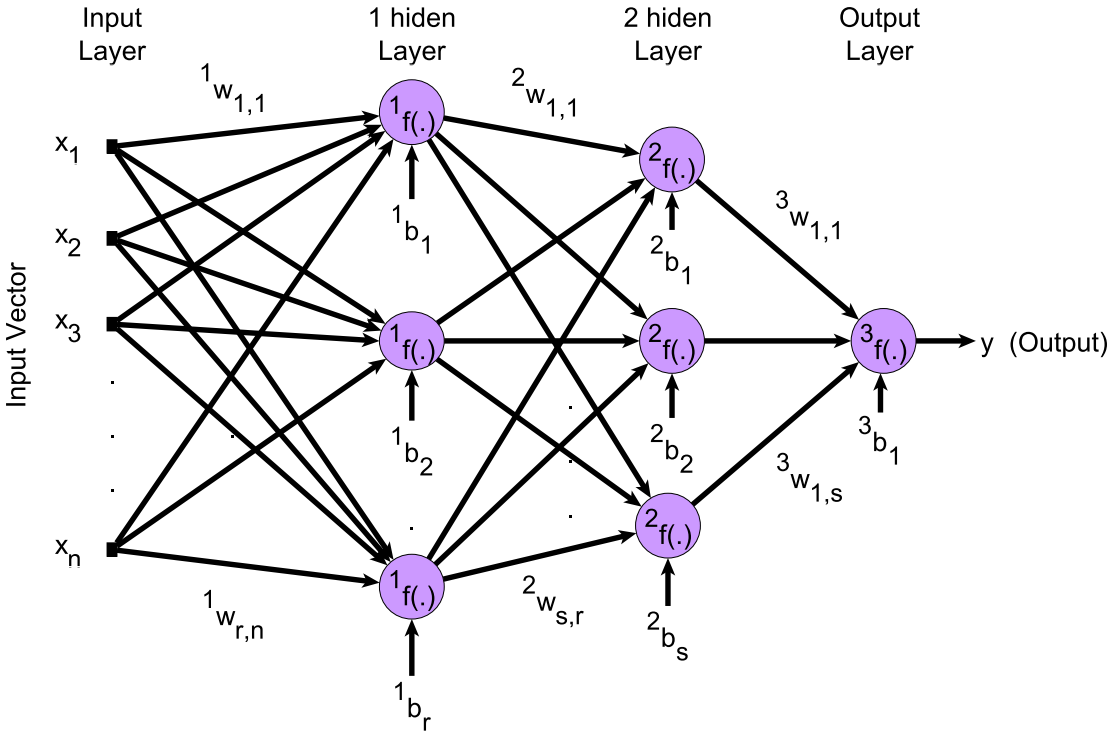

Figure 2.23: Three layers one output ANN

iteratively in a closed loop as shown in Figure 2.24 until the stopping criteria or an acceptable error performance is reached. Example of supervised learning algorithms in-

cludes: backpropagation, Widrow-Hoff rule, Gradient descent, Delta rule, etc. Because of the nature of the problem to be solved in this research, all the ANN prediction models used in this research are trained using a hybrid supervised learning algorithm consisting of meta-heuristic algorithms and backpropagation algorithm Ortega-Zamorano et al. (2015). The meta heuristics are configured to serve as global searchers after which the backpropagation algorithm is use as a local searcher to exploit in detail the decision regions already explored by the metaheuristics algorithms. This has proved to be an efficient way of minimization of premature convergence at local optimum or plateau which is the major problem associated with backpropagation algorithm.

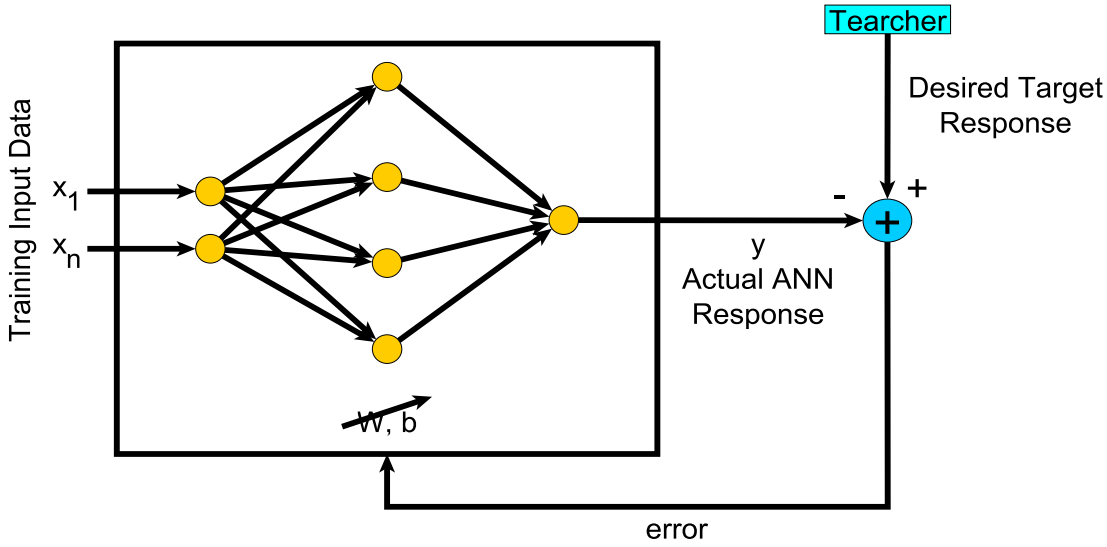

Figure 2.24: Block diagram of supervised learning

2. Unsupervised learning

This is also referred to as self-organized learning where there is no known target outputs or external teacher to oversee the training process. Since there is no desired output, the ANN self adjust its weights to learn and develops the ability to form an internal representation for encoding the features, patterns or trends associated with the input data. This type of learning is often used for data clustering, features extraction and solving classification problems. Example of unsupervised learning include: competitive learning, Hebbian learning and Kohonen's learning rule. The block diagram of unsupervised learning is as shown in Figure 2.25.

3. Hybrid learning This type of learning mechanism consist of both unsupervised and supervised learning as shown in Figure 2.26. It is often used for pattern recognition.

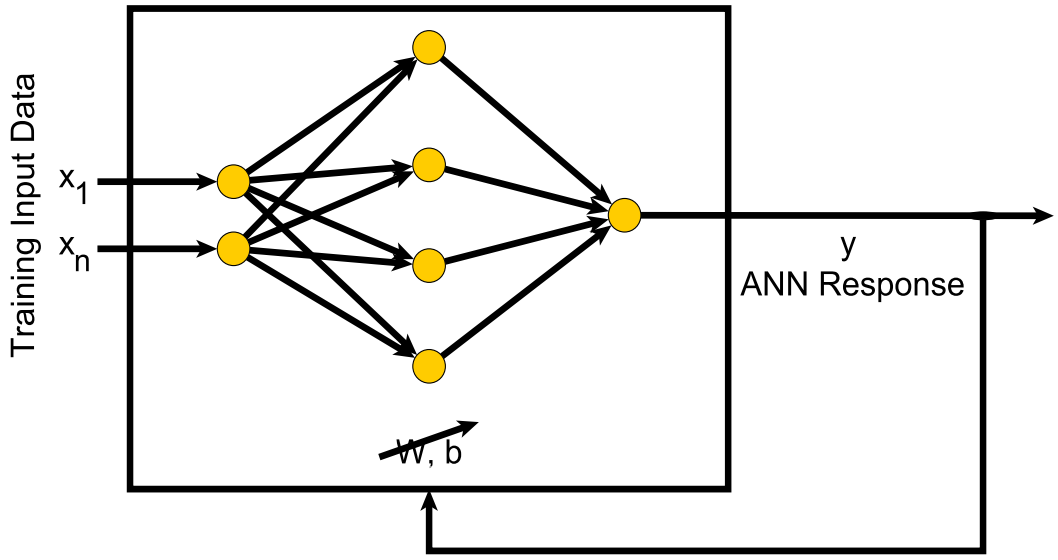

Figure 2.25: Block diagram of unsupervised learning

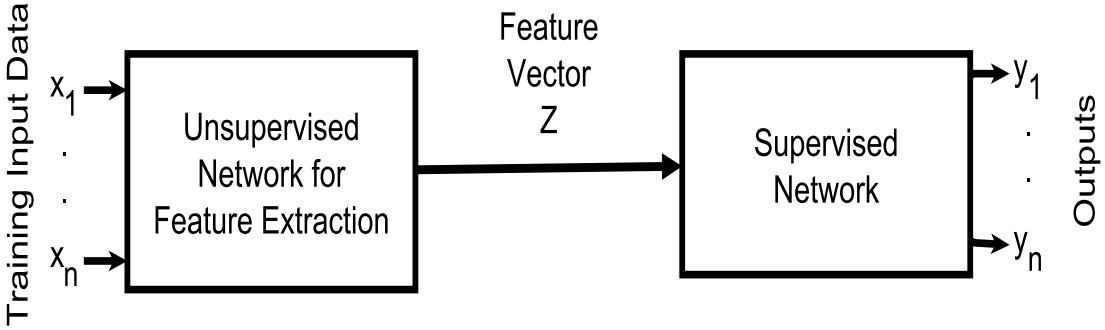

Figure 2.26: Block diagram of hybrid learning ANN

2.5.2 Support Vector Machine

Support Vector Machine (SVM) is a supervised learning system based on statistical theory which has found great application in solving classification and pattern recognition problems. Though the main success of SVM has been in solving pattern-classification problems, recent advances has extended the application of SVM to solving nonlinear-regression problems, Haykin (2008). SVM designed to model regression problem is referred to as support vector regression (SVR). In SVM regression, the input space x is first mapped onto a high m dimensional feature space using certain non-linear transformation. A linear model $f(x, w)$ is then constructed in the feature space which linearly maps the feature space to the output space as shown in Equation (2.47), Haykin (2008). Many time series regression prediction models use certain lost functions during the training phase for minimization of the empirical risk, among these loss functions are mean square error, square error, absolute error, absolute relative percentage error, root mean square error, etc. In SVM regression, a different loss function $L_\varepsilon(y, f(x, w))$

called ε-insensitive loss function is used, Equation (2.48), Vapnik (1999), Vapnik (1998). When the error is within the threshold ε, it is considered as zero, beyond the threshold ε, the loss function (error) is computed as the difference between the actual error and the threshold as depicted in Equation (2.48). The empirical risk function $R_{emp}(w)$ of support vector regression is as shown in Equation (2.49). The goal of SVR model is to approximate an unknown real-value function depicted by Equation (2.46). Where x is a multivariate input vector while y is a scalar output, and δ is independent and identically distributed (i.i.d.) zero mean random noise or error. The model is estimated using finite training samples (x_i, y_i) for $i = 1, \cdots, n$ where n is the number of training samples. For this study, the input vector x of the SVR model consist of past recent RF power, frequency or channel, and current time domain data while the scalar output y is the current RF power in Decibels (dB). Though the training of SVM involved non-linear optimization of the parameters (weights and bias), one of its major advantages is that the objective function is convex, hence the solution of the optimization problem can be easily obtained, Bishop (2006).

$$y = r(x) + \delta \tag{2.46}$$

$$f(x, w) = \sum_{j=1}^{m} w_j g_j(x) + b \tag{2.47}$$

Where $g_j(x),\ j = 1, \cdots, m$ refer to set of non-linear transformations, w_j are the weights and b is the bias.

$$L_\varepsilon(y, f(x, w)) = \begin{cases} 0 & \text{if } |y - f(x, w)| \leqslant \varepsilon \\ |y - f(x, w)| - \varepsilon & \text{otherwise} \end{cases} \tag{2.48}$$

$$R_{emp}(w) = \frac{1}{n} \sum_{i=1}^{n} L_\varepsilon(y_i, f(x_i, w)) \tag{2.49}$$

Support vector regression model is formulated as the minimization of the following objective functions, Haykin (2008):

$$\text{minimise} \quad \frac{1}{2}\|W\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i{}^*) \tag{2.50}$$

$$\text{subject to} \quad \begin{cases} y_i - f(x_i, w) - b \leqslant \varepsilon + \xi_i{}^* \\ f(x_i, w) + b - y_i \leqslant \varepsilon + \xi_i \\ \xi_i, \xi_i{}^* \geqslant 0, \quad i = 1, \cdots, n \end{cases} \tag{2.51}$$

The non-negative constant C is a regularization parameter that determines the trade off between model complexity (flatness) and the extent to which the deviations larger than ε will be tolerated in the optimization formulation. It controls the trade-off between achieving a low training and validation error, and minimizing the norm of the weights. Thus the model generalization is partly dependent on C. The parameter C enforces an upper bound on the norm of the weights, as shown in Equation (2.52). Very small value

of C will lead to large training error while infinite or very large value of C will lead to over-fitting resulting from large number of support vectors, Alpaydin (2004). The slack variables ξ_i and ξ_i^* represent the upper and lower constrains on the output of the system. These slack variables are introduced to estimate the deviation of the training samples from the ε-sensitive zone thus reducing model complexity by minimizing the norms of the weights, and at the same time performing linear regression in the high dimensional feature space using ε-sensitive loss function. The parameter ε controls the width of the ε-insensitive zone used to fit the training data. The number of support vectors used in constructing the support vector regression model (function) is partly dependent on the parameter ε. If ε-value is very large, few support vectors will be selected, on the contrary, bigger ε-value results in a more generalized model (flat estimate). Thus both the complexity and the generalization capability of the network also depend on its value. One other parameter that can affect the generalization and accuracy of a support vector regression model is the kernel parameter and the type of kernel function used as shown in Equation (2.54) to (2.58). SVR optimization problem constitute a dual problem with a solution given by:

$$f(x) = \sum_{i=1}^{s} (\alpha_i - \alpha_i^*) K(x, x_i) + b \tag{2.52}$$

The dual coefficients in equation (2.52) are subject to the constrains $0 \leqslant \alpha_i \leqslant C$ and $0 \leqslant \alpha_i^* \leqslant C$. Where s is the number of support vectors, $K(x, x_i)$ is the kernel function, b is the bias, while α and α^* are Lagrange multipliers. The training samples x_i with non-zero coefficients in Equation (2.52) are called the support vectors. The general expression for the kernel is depicted in Equation (2.53). Any symmetric positive definite function, which satisfies Mercer's conditions can be used as kernel function, Haykin (2008). In this study, five kernel were used, i.e. the Radial Basis Function (RBF), Gaussian Radial Basis Function (GRBF), Exponential Radial Basis Function (ERBF) kernel, Linear Kernel (LK) and Polynomial kernel given by Equations (2.54), (2.55), (2.56), (2.57) and (2.58) respectively, Haykin (2008).

Mercer's theorem for kernel function: This state that:

- The kernel matrix is symmetric positive definite
- Any symmetric positive definite matrix can be regarded as kernel matrix i.e. as an inner product matrix in some space

$$K(x, x_i) = \sum_{j=1}^{m} g_j(x) g_j(x_i) \tag{2.53}$$

- Radial basis function kernel:

$$K(x, x_i) = e^{(-\gamma \|x - x_i\|^2)} \tag{2.54}$$

- Gaussian kernel

$$K(x, x_i) = e^{(-\frac{\|x - x_i\|^2}{2\sigma^2})} \tag{2.55}$$

- Exponential kernel:

$$K(x, x_i) = e^{(-\frac{\|x-x_i\|}{2\sigma^2})} \tag{2.56}$$

- Linear kernel:

$$K(x, x_i) = x^T x_i + c \tag{2.57}$$

- Polynomial kernel:

$$K(x, x_i) = (\alpha x^T x_i + c)^d \tag{2.58}$$

- Multilayer perceptron kernel:

$$K(x, x_i) = \tanh(\rho\langle x, x_i\rangle + \varphi) \tag{2.59}$$

  The multilayer perceptron can only be used as kernel for certain values of the scale parameter ρ and the offset parameter φ which satisfied the Mercer's conditions. In this case the support vectors x_i correspond to the inputs of the first layer while the Lagrange multipliers represent the weights.

- Splines kernel:

$$K(x, x_i) = 1 + \langle x, x_i\rangle + \frac{1}{2}\langle x, x_i\rangle \min(x, x_i) + \frac{1}{6}\min(x, x_i)^3 \tag{2.60}$$

- Additive kernel:

$$K(x, x_i) = \sum_{q=1}^{D} K_q(x, x_i) \tag{2.61}$$

  Since the sum of two or more positive definite functions is also positive definite, more complex kernels can be obtained by summing different kernels. where $q = 1, \cdots, D$ represent D different kernels e.g. with $q = 1$ for linear kernel, $q = 2$ for RBF and so forth.

- Tensor product kernel:

$$K(x, x_i) = \prod_{q=1}^{D} K_q(x, x_i) \tag{2.62}$$

  Just as the addition of positive definite functions is also positive definite, Equation (2.61), so is their products. Hence new kernel can also be formed via tensor product of different kernel, Equation (2.62).

The adjustable constant parameter c of the linear kernel, equation 2.57 was set to 0, the slope α and the constant c of the polynomial kernel were set to 1, Equation (2.58), while the degree of the polynomial d, the kernel parameters σ and γ were evolved using deferential and particle swarm optimization algorithms. The approach adopted for evolving the meta parameter of the SVM regression models are explained in Section 2.7 as part of the contribution of this thesis.

2.6 Optimization Algorithms

A brief description of the metaheuristic optimization algorithms used to optimised the prediction models are presented in this section. The global search capability of population based differential evolutionary and swarm intelligence algorithms are combined with the single solution local search advantages of BPA to evolve the weights and biases of the optimized ANN prediction model as described in Section 4.3.1. The same population based optimization algorithms are used to evolves the metaparameters (C, ε and γ) of the SVR model as detailed in Section 4.3.1.

2.6.1 Differential Evolution Algorithms

Differential evolution (DE) optimization algorithms are population based direct search algorithms used to solve continuous optimization problems, Price et al. (2005), Qin et al. (2009). The DE strategies are simple with only few free metaparameters. The DE scheme was first proposed by Price and Storn, Storn and Price (1997), since then, DE has experienced significant improvements resulting from different mutation strategies. DE aims at evolving NP population of D dimensional vectors which encodes the G generation candidate solutions $X_{i,G} = \left\{X^1_{i,G}, \cdots X^D_{i,G}\right\}$ towards the global optimum, where $i = 1, \cdots, NP$. The initial candidate solutions at $G = 0$ are evolves in such a way as to cover the search space as much as possible by uniformly randomizing the candidates within the decision space using Equation (2.63).

$$X_{i,G} = X_{min} + rand(1,0) \cdot (X_{max} - X_{min}) \tag{2.63}$$

Where NP is the population size (candidate solutions) which remains unaltered through out the evolutionary process, $i = 1 \cdots NP$. X_{min} and X_{max} are the lower and upper bound of the decision space respectively, where $X_{min} = \left\{X^1_{min} \cdots X^D_{min}\right\}$ and $X_{max} = \left\{X^1_{max} \cdots X^D_{max}\right\}$, D is the dimensionality of the candidate solution and $rand(1,0)$ is a uniformly distributed random number between 0 and 1.

Mutation

For every individuals (target vectors) $X_{i,G}$ at generation G, a mutant vector $V_{i,G}$ called the provisional or trial offspring is generated via certain mutation schemes. Some of the commonly used mutation strategies in DE framework are as shown in equations 2.64 to 2.69, Price et al. (2005), Qin et al. (2009).

- DE/rand/1:

$$V_{i,G} = X_{r_1,G} + F \cdot (X_{r_2,G} - X_{r_3,G}) \tag{2.64}$$

- DE/rand/2:

$$V_{i,G} = X_{r_5,G} + F \cdot (X_{r_1,G} - X_{r_2,G}) + F \cdot (X_{r_3,G} - X_{r_4,G}) \tag{2.65}$$

- DE/current-to-rand/1:

$$V_{i,G} = X_{i,G} + K \cdot (X_{r_1,G} - X_{i,G}) + F \cdot (X_{r_2,G} - X_{r_3,G}) \tag{2.66}$$

- DE/best/1:

$$V_{i,G} = X_{best,G} + F \cdot (X_{r_1,G} - X_{r_2,G}) \tag{2.67}$$

- DE/rand-to-best/1:

$$V_{i,G} = X_{i,G} + F \cdot (X_{best,G} - X_{i,G}) + F \cdot (X_{r_1,G} - X_{r_2,G}) \tag{2.68}$$

- DE/best/2

$$V_{i,G} = X_{best,G} + F \cdot (X_{r_1,G} - X_{r_2,G}) + F \cdot (X_{r_3,G} - X_{r_4,G}) \tag{2.69}$$

Where the indexes r_1, r_2, r_3 and r_4 are mutually exclusive positive integers and distinct from i. These indexes are generated at random within the range $[1 \cdots PN]$. $X_{best,G}$ is the individual with the best fitness at generation G while F is the mutation constant which control the amplification of the difference vector e.g. $F \cdot (X_{r_2,G} - X_{r_3,G})$. The mutation schemes that depends on the best candidate solution $X_{best,G}$ as shown in Equations (2.67) to (2.69) with either exponential or binomial crossover often proved to converge faster to a good solution when solving uni-modal problems. Howbeit they are more likely to be trapped at local optimum leading to premature convergence when solving multi-modal problems as they bears strong exploitative capabilities. The mutation strategies that does not involve the use of the best solution e.g. Equation (2.64) often converges slowing with strong explorative move over the search space. As a results, they are more fitted for addressing multi-modal problems. The rotation invariant mutation scheme depicted by Equation (2.66) in which the genes of the mutant $V_{i,G}$ is obtained by mutating the genes of its parent $X_{i,G}$ and other candidate solutions, has been proved to be efficient in solving multi-objective optimization problems.

Crossover

After the mutants are generated, the offspring $U_{i,G}$ are produced by performing a crossover operation between the target vector $X_{i,G}$ and its corresponding provisional offspring $V_{i,G}$. The two crossover schemes i.e. exponential and binomial crossover are used in this study for all the DE algorithms implemented, Zaharie (2007). The binomial crossover copied the jth gene of the mutant vector $V_{i,G}$ to the corresponding gene (element) in the offspring $U_{i,G}$ if $rand(0,1) \leqslant Cr$ or $j = j_{rand}$. Otherwise it is copied from the target vector $X_{i,G}$ (parent). The crossover rate Cr is the probability of selecting the offspring genes from the mutant while j_{rand} is a random number in the range $[1\ D]$, this ensure that at least one of the offspring gene is copied from the mutant. The crossover rate Cr is to be chosen by the user. If Cr is small it will result in exploratory moves parallel to a small number of axes of the decision space .i.e. many of the genes of the

offspring will come from its parent than from the mutants, consequently the offspring will resemble its parent. In this way, the DE will serve as a local searcher as it bear strong exploitative capabilities than being explorative, Montgomery (2010). On the other hand, large values of Cr will lead to moves at angles to the search space's axes as the genes of the offspring are more likely to come from the provisional offspring (mutant vector) than its parent. This will favour explorative moves, Montgomery (2010). The binomial crossover is represented as:

$$U_{i,G}^{j} = \begin{cases} V_{i,G}^{j} & \text{if} \quad (rand(0,1) \leqslant Cr \quad \text{or} \quad j = j_{rand}) \\ X_{i,G}^{j} & \text{otherwise} \end{cases} \tag{2.70}$$

For exponential crossover, the genes of the offspring are inherited from the mutant vector $V_{i,G}$ starting from a randomly selected index j in the range $[1\ D]$ until the first time $rand(0,1) > Cr$ after which all the other genes are inherited from the parent $X_{i,G}$. The exponential crossover is as shown in Algorithm 2.2. The introduction of $j = randi(1, D)$ such that $U_{i,G}^{j} = V_{i,G}^{j}$ in exponential crossover is to ensure that at least one of the genes of the offspring should come from the mutant vector. In this way, the offspring will be different from its parent with at least one gene.

Algorithm 2.2 Exponential Crossover

```
    U_{i,G} = X_{i,G}
2:  generate j = randi(1, D)
    U^j_{i,G} = V^j_{i,G}
4:  k = 1
    while rand(0,1) ⩽ Cr AND k < D do
6:      U^j_{i,G} = V^j_{i,G}
        j = j + 1
8:      if j = D then
            j = 1
10:     end if
        k = k + 1
12: end while
```

Selection process

When the offspring $U_{i,G}$ is birthed via the crossover scheme, to determine whether the offspring should replaced its parent $X_{i,G}$ or not in the next generation, a greedy selection schemes is employed. The cost functions (commonly refered to as the fitness functions) $f(U_{i,G})$ and $f(X_{i,G})$ of the offspring and its parent respectively are computed. If $f(U_{i,G}) \leqslant f(X_{i,G})$ the offspring will replaced its parent in the next generation i.e. $X_{i,G+1} = U_{i,G}$ otherwise its parent will be allowed to continue in the next generation $X_{i,G+1} = X_{i,G}$. The fitness function used in this research is the mean square error (MSE) as explained in Section 4.2.1.

2.6.2 jDE

The jDE scheme is a self-adaptive way of improving DE performance with a very modest programming effort, Brest et al. (2006). There are three control parameters in DE, these are: the mutation constant F, the crossover probability Cr and the pool population NP. These control parameters are problem dependent and bear strong impact on the performance or efficiency of the DE in solving a given optimization problem. Optimum choice of these parameters for a given problem is challenging, and constitute another optimization problem. To address the problems associated with the difficulties in choosing the optimum or near optimum values of F and Cr, the jDE algorithm is proposed Brest et al. (2006), which enhances the pool of DE search moves by including a certain degree of randomization into the original DE framework, Neri and Tirronen (2010). In jDE, the values of mutation and crossover are encoded within each individual candidate solution. For example, the candidate solution X_i will be composed of

$$X_i = (X_i[1], X_i[2], \ldots, X_i[n], F_i, Cr_i)\,. \tag{2.71}$$

Hence, at every generation, the offspring is generated for each individual with the parameters F_i and Cr_i belonging to its parent. Further more, these parameters are periodically updated on the basis of the following randomized criterion:

$$F_i = \begin{cases} F_l + F_u rand_1, & \text{if} \quad rand_2 < \tau_1 \\ F_i, & \text{otherwise} \end{cases} \tag{2.72}$$

$$Cr_i = \begin{cases} rand_3, & \text{if} \quad rand_4 < \tau_2 \\ Cr_i, & \text{otherwise} \end{cases} \tag{2.73}$$

Where $rand_j$, $j \in \{1, 2, 3, 4\}$, are uniform pseudo-random values between 0 and 1; τ_1 and τ_2 are constant values which represent the probabilities of updating the parameters F_i and Cr_i respectively, F_l and F_u are constant values which represent the lower and upper limits of F respectively i.e. $F_l \leqslant F_i \geqslant F_u$.

2.6.3 Super-fit DE

The application of a super-fit logic has proved to be one among the efficient ways of hybridising DE schemes with a local searcher, Caponio et al. (2009). In super-fit logic, a problem is partially solved using a local searcher and the evolves improved solution is injected into the initial DE population (candidate solutions). When a high performance individual is injected into the DE population, the DE is led by it and quickly improved the super-fit thereby producing more fitter candidate solutions. In this research, the initial local search is perform using a Short Distance Exploration algorithm commonly called the S algorithm Caraffini et al. (2013; 2014b;a). The S algorithms is a greedy, single solution local searcher that perform moves along the axes and halves its radius

if unable to obtain a better solution, the pseudocode of the S algorithm is presented in Algorithm 2.3. In this research, 20% of the generation or number of iteration is assigned to the S algorithm while the remaining 80% to the other population based multiple solution meta-heuristic optimization algorithms used e.g. DE and PSO. Without lost of generality, the hyperparameters of the SVR and the weights of the ANN were first evolved using S algorithms and then the solution is injected into the DE population for further search toward the global optimum. For this research, four super-fit variants are implemented, referred to as SDE and SDES for both binomial and exponential crossover. In SDE the problem is partially solved starting with S algorithm and the solution obtained is injected into the DE population where the final best solution is evolved, while in SDES, the best solution evolved by DE is further fine tuned using S algorithm.

Algorithm 2.3 Short Distance Exploration

$x_s = x_0$
$x_{old} = x_0$
for $i = 1 : n$ **do**
  $\delta[i] = \alpha \left(x^U[i] - x^L[i]\right)$ $(\alpha \in]0, 1])$
  $i \leftarrow i + 1$
end for
while *stop condition not met* **do**
  for $i = 1 : n$ **do**
    $x_{trial}[i] = x_s[i] - \delta[i]$
    if $f\left(x_{trial}\right) \leqslant f\left(x_s\right)$ **then**
      $x_s = \mathbf{x_{trial}}$
    else
      $x_{trial}[i] = x_s[i] + \frac{\delta}{2}$
      if $f\left(x_{trial}\right) \leqslant f\left(x_s\right)$ **then**
        $x_s \leftarrow x_{trial}$
      end if
    end if
    $i \leftarrow i + 1$
  end for
  if $f\left(x_s\right) == f\left(x_{old}\right)$ **then**
    $\delta \leftarrow \frac{\delta}{2}$
  end if
end while
Output $\mathbf{x_s}$

Toroidal bound

After the mutation is carried out, some of the genes, dimensions or elements of the mutant vector may fall outside their specified search or decision space $[X^j_{min}\ X^j_{max}]$. It became necessary that before the offspring are form via any of the two crossover schemes, each genes $V^j_{i,G}$ of the provisional offspring (mutant vector) $V_{i,G}$ should be made to be in the range $X^j_{min} \leqslant V^j_{i,G} \leqslant X^j_{max}$, where $V^j_{i,G}$ referred to the j^{th} gene. There are many heuristic empirical proposals on how out-layer genes should be truncated to fall within their decision space. A simple approach which is often used is as shown in Algorithm 2.5, where $V^j_{i,G} = X^j_{max}$ if $V^j_{i,G} > X^j_{max}$ and $V^j_{i,G} = X^j_{min}$ if $V^j_{i,G} < X^j_{min}$ i.e. all out-layer genes will either be assigned X^j_{mim} or X^j_{max}, Liang et al. (2006). One efficient way of handling out-layers is the toroidal bound approach Akutekwe et al. (2014), Iliya et al. (2014d). In this case, instead of constraining all out-layer genes to only two states i.e. X^j_{mim} or X^j_{max}, outl-ayers can take other continuous values within the decision space

Algorithm 2.4 General DE Framework algorithm

**** INITIALIZATION ****

At generation $G = 0$, initialized a population of NP candidate solutions of D dimension uniformly over the decision space

$X_{i,G} = X_{min} + rand(1,0) \cdot (X_{max} - X_{min})$, where $i = 1 \cdots NP$

$X_{min} = \{X^1_{min} \cdots X^D_{min}\}$

$X_{max} = \{X^1_{max} \cdots X^D_{max}\}$

Evaluate the fitness functions $f(X_{i,G})$ of the parents at G=0;

for i=1:NP **do**

  Evaluate $f(X_{i,G})$ for each target vector (parent) $X_{i,G}$

end for

Select initial best candidate $X_{best,G}$ and best fitness $f(X_{best,G})$

while Short of total number of generations or stopping criteria **do**

**** MUTATION ****

For each parent $X_{i,G}$, generate a mutant vector $V_{i,G}$ as:

for i=1:NP **do**

  Generate $V_{i,G} = \left\{V^1_{i,G} \cdots V^D_{i,G}\right\}$ for each $X_{i,G}$ using any of the equations 2.64 to 2.69

end for

Before crossover operation is carried out, each gene (dimension) of the mutant vector is constrained to be within its decision space. For this research a toroidal bound is used as detailed in Figure 2.6

**** CROSSOVER ****

For each parent $X_{i,G}$, generate the offspring $U_{i,G} = \left\{U^1_{i,G} \cdots U^D_{i,G}\right\}$ using either binomial or exponential crossover given by equation 2.70 and Figure 2.2 respectively.

1: Binomial crossover

for i=1:NP **do**

  $j_{rand} = randi(1, D)$

  for j=1:D **do**

$$U^j_{i,G} = \begin{cases} V^j_{i,G} & \text{if} \quad (rand(0,1) \leqslant Cr \quad \text{or} \quad j = j_{rand}) \\ X^j_{i,G} & \text{otherwise} \end{cases}$$

  end for

end for

2: Exponential crossover

$U_{i,G} = X_{i,G}$

generate $j = randi(1, D)$

$U^j_{i,G} = V^j_{i,G}$

$k = randi(1, D)$

while $rand\,(0,1) \leqslant Cr$ AND $k < D$ **do**

  $U^j_{i,G} = V^j_{i,G}$

  $j = j + 1$

  if $j == n$ **then**

    $j = 1$

  end if

  $k = k + 1$

end while

**** SELECTION ****

Evaluate the fitness functions $f(U_{i,G})$ for each offspring and compare it with the fitness function of its parent $f(X_{i,G})$ and the best fitness $f(X_{best,G})$

for i=1:NP **do**

  Evaluate $f(U_{i,G})$

  if $f(U_{i,G}) \leqslant f(X_{i,G})$ **then**

    $f(X_{i,G}) = f(U_{i,G})$

    $X_{i,G} = U_{i,G}$

  end if

  if $f(U_{i,G}) \leqslant f(X_{best,G})$ **then**

    $f(X_{best,G}) = f(U_{i,G})$

    $X_{best,G} = U_{i,G}$

  end if

end for

Increment generation count $G = G + 1$

end while

based on some empirical rule shown in Algorithm 2.6. This introduce another form of randomization into the optimization framework thus making the approach to be more likely suitable for solving both modal and multi-modal optimization problems. The toroidal bound is used to constrain out-layers in all the meta-heuristics optimization algorithms used in this research.

Algorithm 2.5 Two states outlayer genes

if $V_{i,G}^{j} > X_{max}^{j}$ **then**
  $V_{i,G}^{j} = X_{max}^{j}$
else if $V_{i,G}^{j} < X_{min}^{j}$ **then**
  $V_{i,G}^{j} = X_{min}^{j}$
end if

Algorithm 2.6 Toroidal bound algorithm

for i=1:NP **do**
  for j=1:D **do**
    if $X_{i,G}^{j} > X_{max}^{j}$ **then**
      $\delta = X_{i,G}^{j} - X_{max}^{j}$
      $X_{i,G}^{j} = \delta + X_{min}^{j}$
      if $X_{i,G}^{j} > X_{max}^{j}$ **then**
        $X_{i,G}^{j} = X_{max}^{j}$
      end if
    else if $X_{i,G}^{j} < X_{min}^{j}$ **then**
      $\delta = X_{min}^{j} - X_{i,G}^{j}$
      $X_{i,G}^{j} = \delta + X_{min}^{j}$
      if $X_{i,G}^{j} > X_{max}^{j}$ **then**
        $X_{i,G}^{j} = X_{min}^{j}$
      end if
    end if
  end for
end for

2.6.4 Compact Differential Evolution

The population based DE algorithm variants are not memory efficient even though they are often more robust and accurate in solving many optimization problems than the compact version. The compact version of DE which can easily be implemented in embedded systems such as CRs with limited memory constrained is also used in this study. Compact differential evolution (cDE) algorithm is achieved by incorporating the update logic of real value compact genetic algorithm (rcGA) within DE frame work, Mininno et al. (2011b), Harik et al. (1999), Ahn and Ramakrishna (2003). The steps involves in cDE is as follows: A (2 x n) probability vector PV consisting of the mean μ and standard deviation σ is generated. where n is the dimensionality of the problem (in this research it referred to the number of weights and biases, and SVR meta parameters). At initialization, μ was set to 0 while σ was set to a very large value 10, in order to simulate a uniform distribution. A solution called the elite is sampled from the PV. At each generation (step) other candidate solutions are sampled from the PV according to the mutation schemes adopted as described in Section 2.6.1, e.g. for DE/rand/1 three

candidate solutions X_{r_1}, X_{r_2} and X_{r_3} are sampled. Without lost of generality, each designed variable $X_{r_1}[i]$ belonging to a candidate solution X_{r_1}, is obtained from the PV as follows: For each dimension indexed by i, a truncated Gaussian probability density function (PDF) with mean $\mu[i]$ and standard deviation $\sigma[i]$ is assigned. The truncated PDF is defined by equation (2.74). The CDF of the truncated PDF is obtained. A random number $rand(0,1)$ is sampled from a uniform distribution. $X_{r_1}[i]$ is obtained by applying the random number $rand(0,1)$ generated to the inverse function of the CDF. Since both the PDF and CDF are truncated or normalized within the range $[-1,1]$; the actual value of $X_{r_1}[i]$ within the true decision space of [a, b] is obtain as $(X_{r_1}[i]+1)\frac{(b-a)}{2}+a$. The mutant (provisional offspring) is now generated using the mutation schemes. The offspring is evolved by performing a crossover operation between the elite and the provisional offspring as described in Section 2.6.1. The fitness value of the offspring is computed and compare with that of the elite. If the offspring outperform the elite, it replaces the elite and declared the winner while the elite the loser; otherwise the elite is maintained and declared the winner while the offspring the loser. In this study, the fitness function is the MSE obtain using the test data. The weights and the biases of the ANN and SVR parameters are initialized with the offspring and the MSE is obtain, this is repeated using the elite. The one with the least MSE is the winner. The PV is updated using Equations (2.75) and (2.76). Hence in cDE, instead of having a population of individuals (candidates solutions) for every generation as in normal DE, the population are represented by their probability distribution function (i.e. their statistics), thus minimizing the computational complexity, amount of memory needed, and the optimization time. cDE consist of only one candidate solution called the elite, which is either maintained or replaced in the next generation subject to its fitness. The psuedocode of cDE is as shown in Algorithm 2.7, Mininno et al. (2011b).

$$PDF(\mu[i],\sigma[i]) = \frac{e^{\frac{-(x-\mu[i])^2}{2\sigma[i]^2}}\sqrt{\frac{2}{\pi}}}{\sigma[i](erf(\frac{\mu[i]+1}{\sqrt{2}\sigma[i]}) - erf(\frac{\mu[i]-1}{\sqrt{2}\sigma[i]}))} \tag{2.74}$$

$$\mu^{t+1}[i] = \mu^t[i] + \frac{1}{N_P}(Winner[i] - loser[i]) \tag{2.75}$$

$$\sigma^{t+1}[i] = \sqrt{(\sigma^t[i])^2 + \delta[i]^2 + \frac{1}{N_P}(Winner[i]^2 - loser[i]^2)} \tag{2.76}$$

where $\delta[i]^2 = (\mu^t[i])^2 - (\mu^{t+1}[i])^2$, t = steps or generations, N_P is a virtual population and erf is the error function.

2.6.5 Comprehensive Learning Particle Swarm Optimization (CLPSO)

To compare with other algorithms, two algorithms which are not based on DE schemes rather on swarm intelligence framework were considered. These two particle swarm optimization (PSO) variants are: the standard PSO with inertia weight, Shi and Eberhart

Algorithm 2.7 Compact Differential Evolution Pseudocode

generation t=0
** **PV Initialization** **
for $i = 1$ to n **do**
  Initialize $\mu[i] = 0$
  Initialize $\sigma[i] = 10$
end for
Generate the elite by means of **PV**
while buget condition **do**
  ** **Mutation** **
  Generate 3 or more individuals according to the mutation schemes e.g. X_{r_1}, X_{r_2} and X_{r_3} by means of **PV**
  Compute the mutant $V = X_{r_1} + F \cdot (X_{r_2} - X_{r_3})$
  ** **Crossover** **
  $U = V$, where U = offspring
  for $i = 1 : N$ **do**
    Generate rand(0,1)
    if $rand(0, 1) > Cr$ **then**
      $U[i] = elite[i]$
    end if
  end for
  ** **Elite Selcetion** **
  [Winner Loser] = $compete(U, elite)$
  if $U ==$ Winner **then**
    $elite = U$
  end if
  ** **PV Update** **
  for $i = 1 : n$ **do**
    $\mu^{t+1}[i] = \mu^t[i] + \frac{1}{N_P}(Winner[i] - loser[i])$
    $\sigma^{t+1}[i] = \sqrt{(\sigma^t[i])^2 + \delta[i]^2 + \gamma[i]^2}$
    Where: $\delta[i]^2 = (\mu^t[i])^2 - (\mu^{t+1}[i])^2$
    $\gamma[i]^2 = \frac{1}{N_P}(Winner[i]^2 - loser[i]^2)$
  end for
  $t = t + 1$
end while

(1998) and the CLPSO, Liang et al. (2006). PSO emulates the swarm behaviour of which each member of the swarm adapts it search path by learning from its own experience and other members' experiences. The velocity update of PSO and CLPSO are giving by Equations (2.77) and (2.78) respectively while the particle update for both PSO and CLPSO is given by Equation (2.79). In the inertia weighted PSO, each of the particles learn from its local best *pbest* and the global best *gbest* for all dimension. The parameters C_1 and C_2 are the acceleration constants that reflect the weighting of the stochastic acceleration term that pull each particle toward *pbest* and *gbest* respectively. The inertia weight w is used to facilitate both global and local search. Large w facilitate global search while smaller values favoured local search. In this study w was made to decrease exponentially as the generation progresses. This approach facilitate global search within the early stage (generations) and then start to favour local search as the budget (generation) comes to an end. In standard PSO, all particles learn from it own *pbest* and *gbest* at all time and for all dimension. Constraining the social learning aspect to only the *gbest* lead to premature convergence of the original PSO. Since all particles in the swarm learn from the current *gbest* even if the *gbest* is very far from the global optimum. Thus all the particles stand the risk of been attracted to *gbest* and get trapped in a local optimum especially when solving complex multimodal problems with numerous local optimums. To circumvent the problem of premature convergence associated with the standard PSO, a CLPSO was proposed, Liang et al. (2006). In CLPSO, instead of particles learning from its *pbest* and *gbest* for all dimensions, and

for all generations, each element (dimension) of a particle can learn from any other particle's *pbest* including its own *pbest*. The decision on whether a particle's dimension should learn from it own *pbest* or other particles' *pbest* depends on the probability P_c called learning probability. Each particle has its own P_c. For every dimension d of a particle i a random number in the range [0, 1] is generated, if this random number is greater than P_{ci}, the particular dimension will learn from its own *pbest* otherwise it will learn from another particle's *pbest*. To ensure that at least one of the dimension of each particle learn from another particle's *pbest*, if all dimensions happen to learn from its own *pbest*, one dimension is pick at random and two other particles are pick at random from the population, the selected dimension will learn from the corresponding dimension of the particle with the best fitness (*pbest*). In this study, P_{ci} is given by Equation (2.80) base on a given empirical rule, Liang et al. (2006). The pseudocode of CLPSO and initia PSO are shown in Algorithms 2.8 and 2.9 respectively.

$$\begin{aligned} V_i^d \longleftarrow\ & w_i \cdot V_i^d + C_1 \cdot rand1_i^d \cdot (pbest_i^d - X_i^d) \\ & + C_2 \cdot rand2_i^d \cdot (gbest_i^d - X_i^d) \end{aligned} \tag{2.77}$$

$$V_i^d \longleftarrow w_i V_i^d + c \cdot rand_i^d \cdot \left(pbest_{f_i(d)}^d - X_i^d\right) \tag{2.78}$$

Where $pbest_{f_i(d)}^d$ is any particle's *pbest* including particle i *pbest*. $f_i = [f_i(1), f_i(2) \cdots, f_i(D)]$ defined which particles' *pbests* the particle i should follow. D is particle dimension, $rand_i^d$ is a normal distribution random number in the range [0 1], each particle dimension d has its own $rand_i^d$. X referred to particles' positions (potential solutions) while c is the acceleration pull.

$$X_i^d \longleftarrow X_i^d + V_i^d \tag{2.79}$$

$$P_{ci} = 0.05 + 0.45 \frac{e^{\frac{10(i-1)}{SS-1}} - 1}{(e^{10} - 1)} \tag{2.80}$$

where SS is the swarm size (number of particles).

Algorithm 2.8 Comprehensive Learning Particle Swarm Optimization Pseudocode

INITIALIZATION

At generation $G = 0$, initialized a swarm of size N (possible solutions or positions) X_i and velocities V_i of D dimension uniformly over the decision space

$X_i = X_{min} + rand(1,0) \cdot (X_{max} - X_{min})$, where $i = 1 \cdots N$

$V_{max} = 0.3(X_{max} - X_{min})$ and $V_{min} = -V_{max}$

$V_i = V_{min} + rand(1,0) \cdot (V_{max} - V_{min})$,

Set these initial positions as $pbest$ i.e. $X_i = pbest_i$ and evaluate their fitness functions

Evaluate the fitness functions $f(pbest_i)$ of the particles at G=0;

for i=1:N **do**

    Evaluate $f(pbest_i)$ for each particle position $pbest$

end for

Select the particle with the best fitness as global best $gbest$ and global best fitness $f(gbest)$

$flag_i = 0$, set the refreshing gap m and the learning probabilities pc

for k=1:max _iteration **do**

    $w_k = w_{max} - \frac{(w_{max} - w_{min})(k-1)}{max\_iteration - 1}$

    for $i = 1 : N$ **do**

        if $flag_i \geqslant m$ **then**

            $flag_i = 0$

        end if

        VELOCITY UPDATE

        for $d = 1 : D$ **do**

            if $rand(0,1) > pc_i$ **then**

                $V_i^d = w_k * V_i^d + C * rand_i^d(pbest_i^d - X_i^d)$

            else

                $r = randi(1,N)$, $y = randi(1,N)$ select two particles r and y where $r \neq y \neq i$

                if $f(pbest_r) < f(pbest_y)$ **then**

                    $V_i^d = w_k * V_i^d + C * rand_i^d(pbest_r^d - X_i^d)$

                else

                    $V_i^d = w_k * V_i^d + C * rand_i^d(pbest_y^d - X_i^d)$

                end if

            end if

            $V_i^d = min(V_{max}^d, max(V_{min}^d, V_i^d))$, limit the velocity

        end for

        if All dimensions learned from its $pbest_i$ **then**

            Select a particle $z = randi(1,N)$, and dimension $j = randi(1,D)$ where $z \neq i$

            $V_i^j = w_k * V_i^j + C * rand_i^j(pbest_z^j - X_i^j)$

            $V_i^j = min(V_{max}^j, max(V_{min}^j, V_i^j))$, limit the velocity

        end if

        POSITION UPDATE

        $X_i = V_i + X_i$

        $pbest$ **AND** $gbest$ **UPDATE**

        if All the dimensions of X_i are within the decision space i.e. $X_i \in [X_{min}, X_{max}]$ **then**

            Compute fitness of X_i i.e. $f(X_i)$

            if $f(X_i) < f(pbest_i)$ **then**

                $flag_i = 0$

                $pbest_i = X_i$

                $f(pbest_i) = f(X_i)$

                if $f(X_i) < f(gbest_i)$ **then**

                    $gbest_i = X_i$

                    $f(gbest_i) = f(X_i)$

                end if

            else

                $flag_i = flag_i + 1$

            end if

        end if

    end for

end for

Algorithm 2.9 General PSO Pseudocode with initial weight

INITIALIZATION

At generation $G = 0$, initialized a swarm of size N (possible solutions or positions) X_i and velocities V_i of D dimension uniformly over the decision space

$X_i = X_{min} + rand(1,0) \cdot (X_{max} - X_{min})$, where $i = 1 \cdots N$

$V_{max} = 0.3(X_{max} - X_{min})$ and $V_{min} = -V_{max}$

$V_i = V_{min} + rand(1,0) \cdot (V_{max} - V_{min})$,

Set these initial positions as $pbest$ i.e. $X_i = pbest_i$ and evaluate their fitness functions

Evaluate the fitness functions $f(pbest_i)$ of the particles at G=0;

for i=1:N **do**

    Evaluate $f(pbest_i)$ for each particle position $pbest$

end for

Select the particle with the best fitness as global best $gbest$ and global best fitness $f(gbest)$

for k=1:max _iteration **do**

    $w_k = w_{max} - \frac{(w_{max} - w_{min})(k-1)}{max\_iteration - 1}$

    for $i = 1 : N$ **do**

        VELOCITY AND POSITION UPDATE

        for $d = 1 : D$ **do**

            $V_i^d = w_k * V_i^d + C1 * rand1_i^d(pbest_i^d - X_i^d) + C2 * rand2_i^d(gbest_i^d - X_i^d)$

            $V_i^d = min(V_{max}^d, max(V_{min}^d, V_i^d))$, limit the velocity

            $X_i^d = V_i^d + X_i^d$

            $X_i^d = min(X_{max}^d, max(X_{min}^d, X_i^d))$, limit the position

        end for

        $pbest$ **AND** $gbest$ **UPDATE**

        Compute fitness of X_i i.e. $f(X_i)$

        if $f(X_i) < f(pbest_i)$ **then**

            $pbest_i = X_i$

            $f(pbest_i) = f(X_i)$

            if $f(X_i) < f(gbest_i)$ **then**

                $gbest_i = X_i$

                $f(gbest_i) = f(X_i)$

            end if

        end if

    end for

end for

2.7 Selection of meta-parameters of SVR model

There are three meta-parameters or hyperparameters that determine the complexity, generalization capability and accuracy of support vector machine regression model as explained in Section 2.5.2, these are the C Parameter, ε and the kernel parameter γ, Kecman (2001), Vladimir and MA (2002), Wenjian et al. (2003). Optimal selection of these parameters is further complicated due to the fact that they are problem dependent and the performance of the SVR model depends on all the three parameters. There are many proposals how these parameters can be chosen. It has been suggested that these parameters should be selected by users based on the users experience, expertise and a priori knowledge of the problem, Cherkassky and Mulier (1998), Bernhard and J. (2001), Vapnik (1999), Vapnik (1998). This leads to many repeated trial and error attempts before getting the optimums if possible, and it limit the usage to only experts. Another proposal state that C should be tuned through a very wide range of values and the optimal performance assessed using cross validation techniques John and Cristianini (2004), this is computationally expensive and poor time management. Another proposal Davide and Simon (1999), suggested that C should be equal to the range of

the training data. This seem intuitive, but it does not take into account possible effects of outliers resulting from noisy training data. James and Tsang (2003) and Smola et al. (1998) proposed asymptotically optimal ε-value proportional to noise variance, in agreement with general sources on SVM. One drawback of such proposals is that they do not take into account the effect of training sample size on ε-value, i.e. the lager the training sample size, the smaller the ε-value, Cherkassky and Ma (2004). In this study, C and ε-value are obtain by adopting the empirical formula given by Equations (2.81) and (2.82), Cherkassky and Ma (2004), for obtaining a rough estimate of ε-value and C parameter respectively. After which the ε-value and C parameter are fine tune towards the global optimum using differential and swarm intelligence optimization algorithms. The entire SVM regression model is trained using the proposed hybrid ad-hoc nested algorithm whose outer layer is metaheuristic optimization algorithms variants for sampling the metaparameters (C, ε-value and γ) while the inner layer is convex optimization algorithm for evolving the weights w and bias b of the SVM, Iliya et al. (2015a), Iliya and Neri (2016).

$$C = |\mu| + \tau_c \sigma \tag{2.81}$$

$$\varepsilon = \tau_\varepsilon \sigma \sqrt{\frac{\log n}{n}} \tag{2.82}$$

Where μ and σ are the mean and standard deviation of the training data respectively, n is the number of training samples, the constants τ_c and τ_ε are problem domain specific for optimum C and ε parameters. In order to get the optimum or near optimum values of C and ε, the following empirical rule was used: The minimum and maximum values for τ_c and τ_ε were set. These two limits were obtained after some manual tuning and repeated trials. These values were used to obtain the minimum and maximum values of C and ε thus forming a decision space. Particle swarm and differential evolutionary optimization algorithm variants were used to evolve the optimum or near optimum values of C, ε and γ within the decision (search) space. This approach constrains the search space, to a manageable range leading to fast convergence instead of using very large search space as earlier proposed, John and Cristianini (2004). This approach also captured the effect of outliers resulting from training data which was not captured in any of the proposals, Davide and Simon (1999). The parameters C and ε are multi-optimal and suboptimal space variables Cherkassky and Ma (2004), i.e. for a given problem, there exists more than one pair of C and ε with optimum or near optimum performance, and many local optimums. Therefore varying τ_c and τ_ε manually as suggested Cherkassky and Ma (2004), may result into selection of C and ε at local optimum. To circumvent this deficiency, population based randomization optimization algorithms with a high probability of escaping local optimum were used to evolve C, ε and γ thus resulting into optimum or near optimum values of C, ε and γ, Iliya et al. (2015a), Iliya and Neri (2016).

2.8 Summary

This chapter presents the various spectrum sensing techniques with their merits and demerits. The background theory governing the prediction models and the optimization algorithms used, were discussed and a brief description of the improvements made on the existing algorithms. The next chapter presents spectrum occupancy survey using energy detector. This aims at evaluating the degree of utilization of some channels with the overall aim of identifying underutilised channels that can serve as potential channels to be exploited for cognitive radio applications.

Chapter 3

Spectrum Occupancy Survey

3.1 Introduction

The success of Cognitive Radio (CR) network is largely dependent on the overall world knowledge of spectrum utilization in both time, frequency and spatial domains. This research explored only two varying domains i.e. time and frequency while the spatial domain is kept constant as the experiments were conducted in one geographical location within the same altitude. The majority of current radio systems are not fully aware of their radio spectrum environment and operate within a confirmed specific frequency bands using dedicated static access systems. The current static spectrum allocation policy has resulted in most of the licensed spectra been grossly under utilized while some spectrum are congested leading to low quality of service. The problem of spectrum scarcity and under utilization, coupled with USA Federal Communication Commission (FCC) Communication (2003), approval of some licensed spectrum to be exploited by cognitive radios for efficient spectrum utilization, has motivated many researchers in different parts of the world to conduct spectrum occupancy survey, Bacchus et al. (2010), Qaraqe et al. (2009), Contreras et al. (2011a), Mehdawi et al. (2013), Islam et al. (2008), McHenry et al. (2006). There are two major reasons why spectrum occupancy survey is vital to the implementation of cognitive radio network. The first reason for the survey is to validate the fact that the current spectrum scarcity and underutilization are not due to the fundamental lack of spectrum rather is the direct consequence of current spectrum allocation policy. This will facilitate the implementation of a new paradigm of wireless communication based on Dynamic Spectrum Access (DSA) which is the bedrock of cognitive radio network. Secondly, the spectrum occupancy survey is aimed at determining underutilized bands in both time, frequency and spatial domains that can serve as potential candidate bands to be exploited by cognitive radios for efficient spectrum utilization, provision of higher bandwidth services and minimising the need for centralized spectrum management, OFCOM (2007). In addition to the aforementioned, the spectrum data obtained from the survey can be used to extract patterns (trends) of radio frequency (RF) transmission activities over a wide range of frequency, time and geographical location which can be used to develop both supervised and unsupervised

computational intelligence prediction models for future forecasting. The outcome of the spectrum occupancy survey can be used for interference avoidance analysis, Datla (2004). The importance of the spectrum occupancy survey cannot be over emphasized as the information acquired during the survey can serve as a vital tool for implementing efficient spectrum sharing and spectrum mobility algorithms, Wang (2009). The spectrum occupancy or utilization of some communication bands in Leicester UK, ranging from 50MHz to 860MHz consisting of radio and TV broadcasting channels; and 868MHz to 2.5GHz with emphases on ISM 868 band, GSM900 and GSM1800 bands and ISM 2.4GHz band are presented in this chapter. This study will add to the existing spectrum occupancy survey campaigns conducted in other part of the world in order to facilitate the implementation of cognitive radio network. The spectrum occupancy of the bands within the studied location were modelled using computational intelligent techniques based on Artificial Neural Network (ANN) and Support Vector Machine (SVM) framework as discussed in Section 4.2. This will enable the CR to predictively exploit bands that are more likely to be free at a particular time instant within a given geographical location.

The rest of the chapter is organized as follows: A brief description of the services within the bands considered are presented in Section 3.2 while the experimental set-up and measurements conducted are detailed in Section 3.3. The approached adopted in estimating the spectrum utilization of the bands is discussed in Section 3.4. The results of the spectrum occupancy are presented in Section 3.5 while Section 3.6 gives a brief summary of the findings.

3.2 Services within the bands considered

The spectrum occupancy survey conducted in this study covered a wide range of frequencies ranging from 50 - 2500MHz. Some of the services within the bands considered based on UK frequency allocation policy are as follows:

- 50 - 110 MHz: This band is used for fixed/land mobile communication, amateur, aeronautical radio navigation, radio astronomy, broadcast, and FM sound broadcasting providing both local and national VHF services.
- 110 - 250 MHz: This band is dedicated for used by meteorological satellite, Mobile satellite, space research, space operation, armature satellite, aeronautical mobile, land mobile, maritime mobile, astronomical mobile, broadcasting, radio location, programme making and special events (PMSE).
- 250 - 470 MHz: The services here consist of Mobile satellite, space operation, space research radio astronomy, aeronautical radio navigation, meteorological aids, Met satellite, Radio navigation satellite, armature satellite, fix mobile and scanning telemetry.

- 470 - 600 MHz: The services in this band includes: Broadcasting, TV band IV, land mobile, fixed mobile, and aeronautical radio navigation.
- 600 - 730 MHz: Services in this band include TV broadcast band V, and some PMSE
- 730 - 860 MHz: Services within this band includes: TV Broadcasting, PMSE, fixed mobile satellite, radiolocation, meteorological satellite, earth exploration services, space research, and aeronautical radio navigation.
- 868 - 890 MHz: This constitutes the 868 MHz ISM band which is also used by the government for radio-location and aeronautical radio-navigation.
- 890 - 960 MHz: This is used for fixed and land mobile communication, GSM 900 band services, and Radio location.
- 1700- 1880 MHZ: This is used for fixed and land mobile, GSM 900 band services,meteorological satellite and radio astronomy.
- 2400- 2500 MHz: This band consist of the 2.4 GHz ISM band, also used for fixed and mobile satellite communication, PMSE and radio location services.

3.3 Measurement Setup

The datasets used in this study were obtained by capturing real world RF signals using universal software radio peripheral 1 (USRP 1) for a period of six months within three years. The USRP are computer hosted software-defined radios with one motherboard and interchangeable daughter board modules for various ranges of frequencies. The USRP provides a flexible and powerful platform for testing and implementing software radio systems. The daughter board modules serve as the RF front end. Two daughter boards, SBX and Tuner 4937 DI5 3X7901, having continuous frequency ranges of 400MHz to 4.4GHz and 50 MHz to 860 MHz respectively, were used in this research. The daughter board perform analog operations such as up/down-conversion, filtering, and other signal conditioning. Furthermore, the daughter board serve as a wideband interface to the analog to digital converter (ADC) and digital to analog converter (DAC) of the USRP. The motherboard perform the functions of clock generation and synchronization, analog to digital conversion, digital to analog conversion, host processor interface, and power control. It also decimate the signal to a lower sampling rate that can easily be transmitted to the host computer through a high-speed USB cable where the signal is processed by software. This survey is focused on the following frequencies: 50 - 860 MHz, 868 - 960 MHz, 1.7 - 1.88 GHz and 2.4 - 2.5 GHz bands. The TV channels bandwidth is 8 MHz while GSM 900, GSM 1800, and FM bands has a channel bandwidth of 200 KHz, and ISM 868 have the least channels bandwidth of 25 KHz. The spectrum within the bands considered where divided into subchannels called Physical Resource Block (PRB), each consisting of 300 KHz bandwidth. Without loss

of generality, to capture the RF signal in a given band, the capturing was constrained within one physical resource block at a time using a channel filter; this gives a high frequency resolution at lower sampling rate that can easily be processed by the host computer. But the higher the frequency resolution, the lower the temporal (time) resolution. Another advantage of dividing the bands into small PRB is that the estimated spectrum occupancy will be more accurate since the duty cycle of a band or channel will be equal to the percentage of active PRBs within the band instead of generalising it over the entire channel which is the common practice, Mehdawi et al. (2013), Islam et al. (2008), McHenry et al. (2006). The fact that a channel is busy does not implies that all the PRBs within the channel are active or utilized, hence segmenting the channels into small PRBs will give a more reliable real time information on how efficient the channel is utilized. To ensure that no spectral information is lost, a sampling frequency of 1 MHz is use and 512 samples are obtained for each sample time. This gives a very high frequency resolution of 1.172 KHz since for every 300 KHz physical frequency resource block, there are 256 frequency bins. Since the least channel bandwidth considered in this survey is 25 KHz i.e. for ISM 868 MHz band, frequency resolution of 1.172 KHz is very adequate to capture all the carriers spectral information within the bands. The power was obtained using both the time and frequency domain data. For the frequency domain, after passing the signal through the channel filter, the signal was windowed using hamming window in order to reduce spectral leakage, Fang et al. (2004), Podder et al. (2014). The stream of the data was converted to a vector and decimated to a lower sampling rate that can easily be processed by the host computer at run time. This is then converted to the frequency domain and the magnitudes of the bins were passed to a probe sink. The choice of probe sink is essential because it can only hold the current data and does not increase thereby preventing stack overflow or segmentation fault from occurring. This allows Python to grab the data at run time for further analysis. The interval of time between consecutive sample data was selected at a random value between 5 seconds and 30 seconds. The choice of this range is based on the assumption that for any TV programme, FM broadcast or GSM calls, will last for not less than 5 to 30 seconds. In order to capture all possible trends, the time between consecutive sample data is selected at random within the given range instead of using regular intervals. The experimental setting is as shown in Figure 3.1. The RF power of some of the bands captured during this survey are presented in Figures 3.2, 3.3 and 3.4. The experiment was conducted indoor in the Centre for Computational Intelligence, De Montfort University, Gateway House, Leicester, United Kingdom (GPS location: latitude 52.629472^0N, longitude 1.138028^0W). The signal processing were implemented using GNU-radio which is a combination of C++ for designing the signal processing blocks, and Python for scripting.

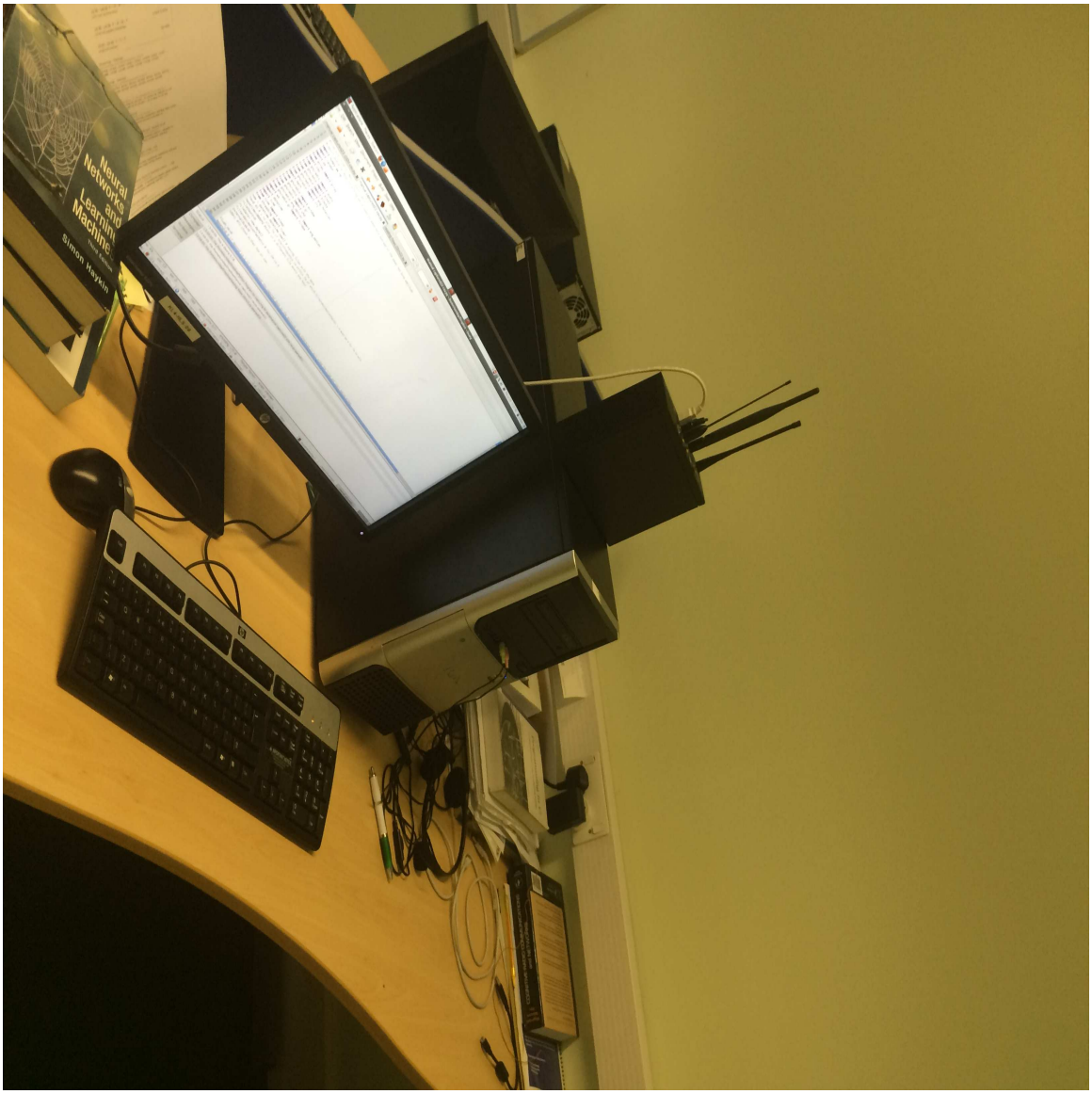

Figure 3.1: Experimental setting

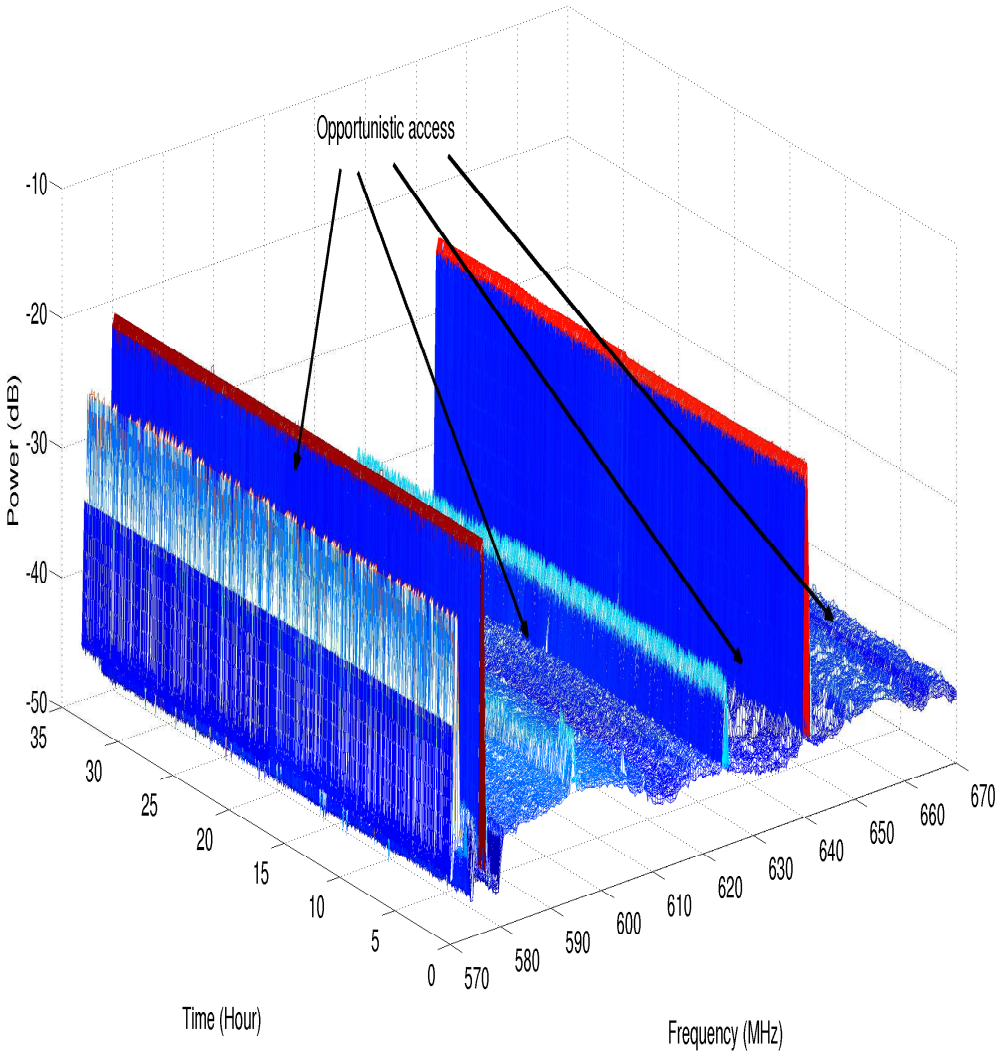

Figure 3.2: TV Band 570 to 670MHz Spectrum

The spectrum captured for three days from 31 May 2015 to 3 June 2015

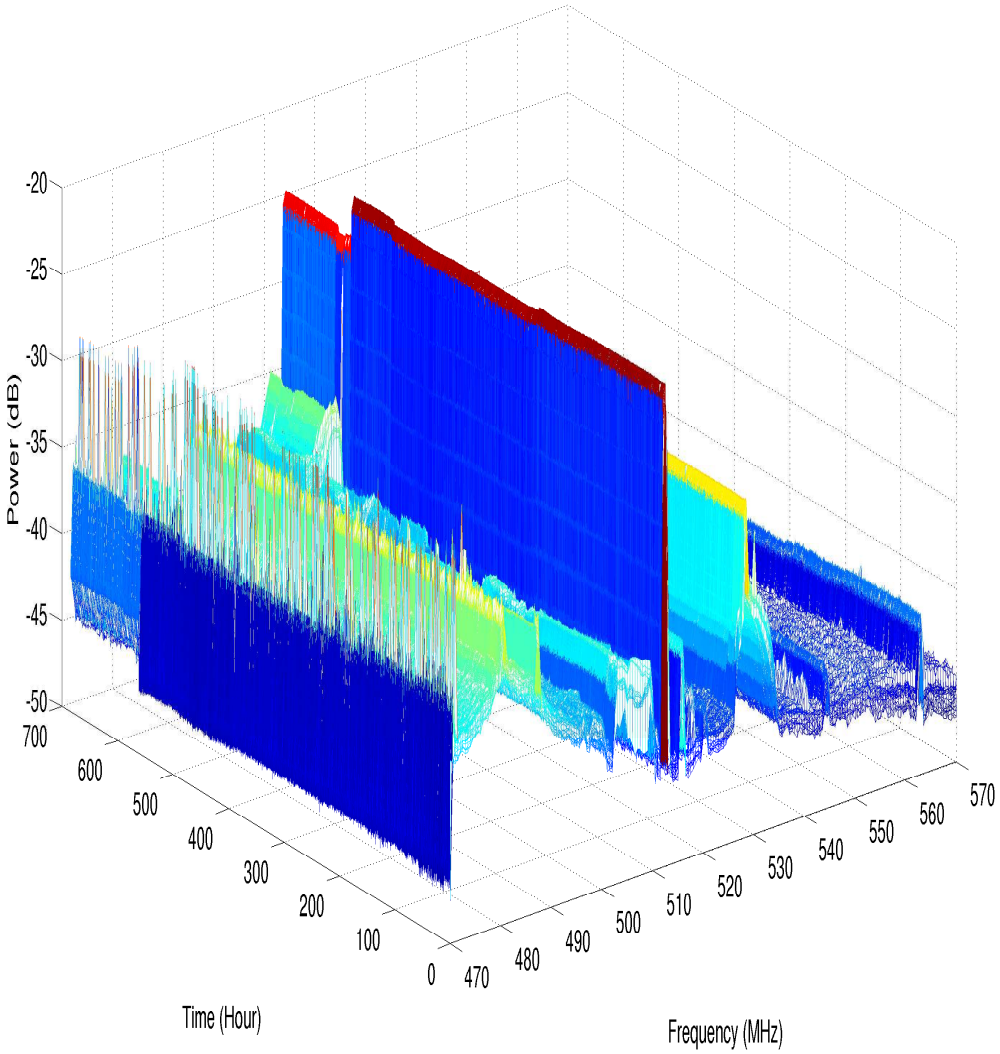

Figure 3.3: TV Band 470 to 570MHz Spectrum

The spectrum is captured within a period of one month i.e. from 27 Sep 2013 18:21 to 26 Oct 2013 06:42

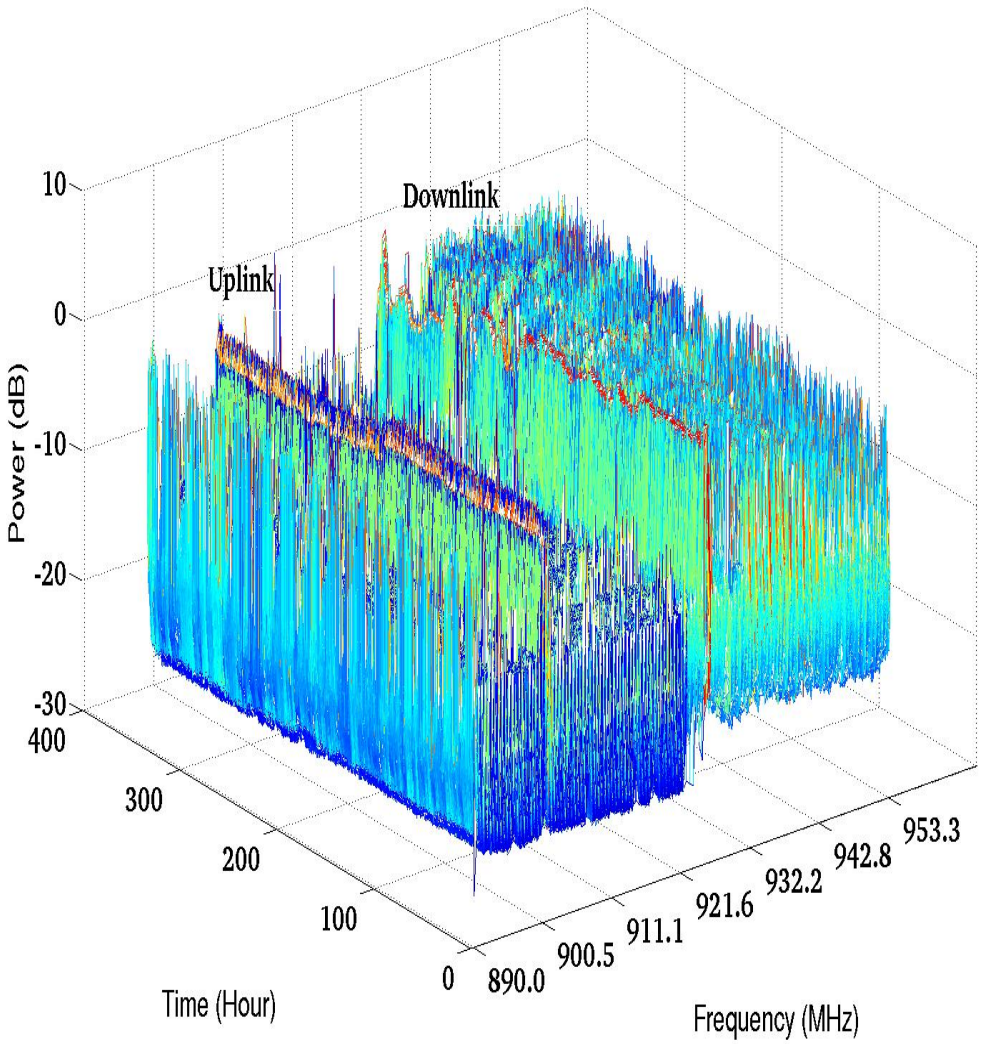

Figure 3.4: GSM 900 band

The GSM 900 band spectrum was captured continuously for sixteen days

3.4 Spectrum Occupancy and Utilization Estimate

One of the major challenges of CR network implementation is real time spectrum sensing, in which the CR senses the primary user (PU) occupancy status (RF activity) Wyglinski et al. (2009). Based on the sensing results the channels are classified using a binary classifier as busy or idle. There are many detection methods that has been proposed, among these are: cooperative sensing, match filter detector, the energy detector, cyclostationary features detector, eigenvalue detector, wavelet transform based estimation and Multitaper spectrum estimation Subhedar and Birajdar (2011b). Of all the proposed methods of spectrum sensing techniques, the energy detector is the only blind method that does not require any a priory knowledge of the PU signal information, hence it is the method used in this survey. The major challenge of energy detection is the correct choice of threshold such that if the energy is above the threshold the channel is considered busy otherwise it is idle (vacant). If the threshold is too low, there will be high probability of false alarm (false positive) i.e. declaring a channel to be busy while it is idle (free). This is safe to the PU but is a waste of communication opportunity to the secondary user (SU) or CR leading to low spectrum utilization. On the other hand, if the threshold is too high, weak PU signals will not be detected resulting to high probability of miss detection (false negative) i.e. declaring a channel to be idled while it is busy. This implies high interference to the PU and lost of communication or (data) to the CR. Although the optimum choice of threshold is a current active research area, there exist an empirical method recommended by International Telecommunication Union (ITU), Bureau (2002). The ITU recommended that the threshold should be 10dB above the ambient noise. It has been proved experimental that a threshold of 10dB above the ambient noise will yield a very low probability of false alarm of about 0.0005% in must cases, Patil et al. (2011). The ambient noise is not flat even for the same band, it changes with frequency, time and geographical location. Based on this empirical recommendation, the optimum choice of the threshold hold depends on the relative cost of false alarm and miss detection. In order to increase the accuracy of the spectrum occupancy estimate; instead of setting a constant threshold over an entire band as adopted by Mehdawi et al. (2013), Islam et al. (2008), McHenry et al. (2006), each frequency resource block (300KHz) has its own threshold. Thus for this study, each frequency bin has its own decision threshold. The ambient noise floors are obtained by replacing the antenna with 50Ω smart antenna terminator as recommended by ITU, and the average ambient noise for each resource block captured within a period of 15 hours were used as the noise floors. In this way the threshold $\top(f)$ varies with frequency since the noise floor $\varrho(f)$ is frequency dependent. To ensure that the threshold meet up with the standard criteria set by ITU, the margin $M(f)$ was set at constant value of 10dB taken into cognition the gain introduced in the codes that run the USRP. The threshold is giving by Equation (3.1).

$$\top(f) = \varrho(f) + M(f) \tag{3.1}$$

Where $\top(f)$ is the threshold of a physical frequency resource block f, $\varrho(f)$ is the average ambient noise power of the given resource block f while $M(f)$ is a constant margin whose values depend on the desired probability of false alarm.

The metric for quantifying the channel or spectrum occupancy (utilization) level is the duty cycle. The duty cycle is defined as the percentage of time the channel is busy. This is the probability that the RF power of a given channel or physical resource block is greater than a predefined threshold. If the i-th power measurement at frequency f (physical frequency resource block) is $P_i(f)$ and the threshold is $\top(f)$, the instantaneous spectrum occupancy rate $B_i(f)$ for frequency f at the i-th measurement is given by Equation (3.2).

$$B_i(f) = \begin{cases} 1 & \text{if } P_i(f) > \top(f) \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

Where i is time index, 1 and 0 are binary status for busy and idle (vacant) respectively. For this spectrum occupancy campaign, the threshold $\top(f)$ is the threshold of resource block f.

The instantaneous spectrum occupancy rate $B_i(f)$ can be modelled as a binary hypothesis test problem with two possible hypotheses H_0 and H_1 for vacant (idle or free) and busy states respectively as depicted in Equation (3.4) Arora and Mahajan (2014). The null hypothesis is H_0 while the alternate hypothesis is H_1, these are assigned binary states 0 and 1 respectively. The phenomenon can be explained further by considering Equation (3.3). In the present of primary user (PU) transmitter signal $x(n)$, the received signal $y(n)$ will contain the amplified version of the PU signal $hx(n)$ and the channel noise $w(n)$ while in the absent of licensed owner i.e. PU, the received signal $y(n)$ contain only noise $w(n)$. Hence the received signal strength is expected to be high when the PU signal $x(n)$ is present as compared to the signal strength when the PU is not active (idle).

$$\begin{aligned} H_0 : &\quad y(n) = w(n) \\ H_1 : &\quad y(n) = hx(n) + w(n) \end{aligned} \tag{3.3}$$

where $w(n)$ is the additive white Gaussian noise, $x(n)$ is the PU signal at time n, h is the channel gain, $y(n)$ is the measured signal while H_0 and H_1 are the two hypothesis for absent and present of PU respectively.

$$B_i(f) = \begin{cases} H_1 & \text{if } P_i(f) > \top(f) \\ H_0 & \text{otherwise} \end{cases} \tag{3.4}$$

Where $B_i(f)$ is the instantaneous spectrum occupancy rate, i is time index, H_1 and H_0 are the two hypotheses for busy and idle (vacant) respectively. $\top(f)$ is the threshold of physical resource block f and $P_i(f)$ is the RF power of f.

The probability of detection and false alarm can be expressed in terms of these two

hypotheses as:

$$Pd_i(f) = Pr\{B_i(f) = 1|H_1\} = Pr\{P_i(f) > \top(f)|H_1\} \tag{3.5}$$

$$Pf_i(f) = Pr\{B_i(f) = 1|H_0\} = Pr\{P_i(f) > \top(f)|H_0\} \tag{3.6}$$

where $Pd_i(f)$, $Pf_i(f)$ and $P_i(f)$ are the probability of detection, probability of false alarm and RF power respectively, of the i observation of resource block f. $\top(f)$ is the decision threshold. From Equations (3.5) and (3.6) it is obvious that the probability of detection and false alarm depends only on the decision threshold $\top(f)$. The probability of miss detection $Pm_i(f)$ can be expressed using Equations (3.7) and (3.8).

$$Pm_i(f) = 1 - Pd_i(f) \tag{3.7}$$

$$Pm_i(f) = Pr\{B_i(f) = 0|H_1\} = Pr\{P_i(f) \leqslant \top(f)|H_1\} \tag{3.8}$$

The channel or resource block duty cycle is the average of the instantaneous occupancy rate of the channel over the entire period of measurement (observation) as depicted by Equation (3.9).

$$\psi(f) = \frac{\sum_{i=1}^{n} B_i(f)}{n} \tag{3.9}$$

Where $\psi(f)$ is the duty cycle of channel or resource block f while n is the total number of time slots the channel is sampled or observed. The duty cycle of the band spectrum is the average of the duty cycles of the channels or resource blocks within the band as given by Equation (3.10). This gives the estimate of the spectrum occupancy (utilization) of the band. From the aforementioned, the maximum attainable duty cycle is 1, thus the product of the duty cycle of a band and the frequency range of the band (bandwidth) gives the estimate of utilization level of the band in Hz. The larger the value of the duty cycle of a band, the higher the occupancy or utilization level of the band.

$$\Phi(b) = \frac{\sum_{i=1}^{N} \psi_i(f)}{N} \tag{3.10}$$

Where $\Phi(b)$ is the duty cycle of band b and N is the total number of channels or physical resource blocks whose power measurements were taken in the band.

3.5 Results of spectrum occupancy survey conducted

In order to capture as much as possible the trends of RF activities within the studied location, the RF signals were captured at different interval of time within a period of three years. The results of some of the spectrum occupancy survey conducted are presented as follows:

Spectrum Occupancy survey for three days

Table 3.1 gives the duty cycle of the bands captured for three days from 31 May to 3 June 2015 while Figures 3.5 to 3.12 depicts the utilization level (occupancy estimate) of each of the channels within the bands. Band 50 to 110 MHz assigned to fixed/land mobile, amateur, aeronautical radio navigation, radio astronomy, broadcast, and FM services has a duty cycle of 3.38%. Only one channel (107.5MHz) within the FM band is utilized apart from possible low power devices operating within the band as shown in Figure 3.5. The band assigned to meteorological satellite, mobile satellite, space research, aeronautical mobile, land mobile, maritime mobile, astronomical mobile, broadcasting, Radio location, space operation, radio astronomy and meteorological aids ranging from 110 to 470 MHz was divided into two bands 110 - 250 MHz and 250 - 470 MHz, each of these two bands was found to have an estimated duty cycle of 4.29% and 3.46% respectively. Thus from 50-470 MHz, the average duty cycle is 3.71%, i.e. the averaged occupancy is 15.582 MHz of the 420 MHz bandwidth. The TV band 470-860 MHz within which there are other services such as aeronautical radio navigation, programme making and special events (PMSE) applications and mobile communication; was divided into three bands i.e. 470-600 MHz, 600-730 MHz and 730-860 MHz. Band 470-600 MHz with a duty cycle of 29.52% is much more utilized than the other two TV bands, Figure 3.8. Band 600-730 MHz is grossly underutilized with an estimated duty cycle of 2.55% while that of band 730-860 MHz is 15.02% as shown in Figure 3.9 and Figure 3.10 respectively. The average occupancy estimate of band 470 - 860 MHz is 15.70% which is equivalent to 61.23 MHz of the total bandwidth 390 MHz. GSM 1800 band appear to be much more utilized than GSM 900 band as shown in Figure 3.12 and 3.11. GSM 1800 band has the highest duty cycle of 51.57% followed by GSM 900 band with 40.85%. The duty cycle of GSM 900 and GSM 1800 bands shown in subplot 3 of Figure 3.11 and Figure 3.12 exhibit a different trend with no frequency resource block having a duty cycle of zero, contrary to those of Figures 3.5 to 3.10 where there are many frequency resource blocks (spectrum) with zero duty cycle for the whole period of measurements (observations). Within the studied location, the ISM 2.4 GHz band has a higher spectrum occupancy with a duty cycle of 18.55% than ISM 868 MHz which has a duty cycle of 14.71%. The summary of the spectrum occupancy results presented in Table 3.1 is graphically depicted in Figure 3.13. From Figure 3.13 the overage occupancy estimate of the bands considered within 50 - 2500 MHz is 18.34%. This implies that out of the 1182 MHz total bandwidth considered, the overage spectrum often in used is 216.78 MHz which is grossly inefficient. In addition to this, from the spectrum occupancy survey results presented in Table 3.1 and Figures 3.5 to 3.12, it is obvious that based on the International Telecommunication Union (ITU) recommended standard, some bands are grossly underutilized. This calls for a new paradigm of spectrum access for efficient spectrum utilization. This survey confirmed the fact that the current spectrum scarcity is not the fundamental lack of spectrum, rather is the direct consequence of static spectrum allocation policy.

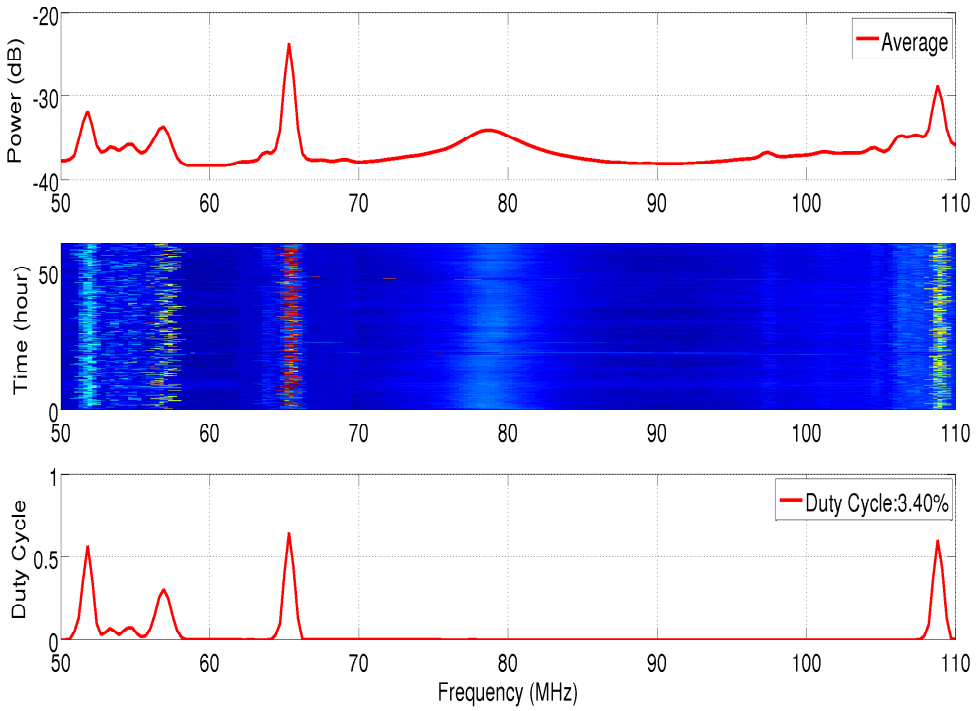

Figure 3.5: Duty cycle and spectrogram of 50 to 110 MHz band

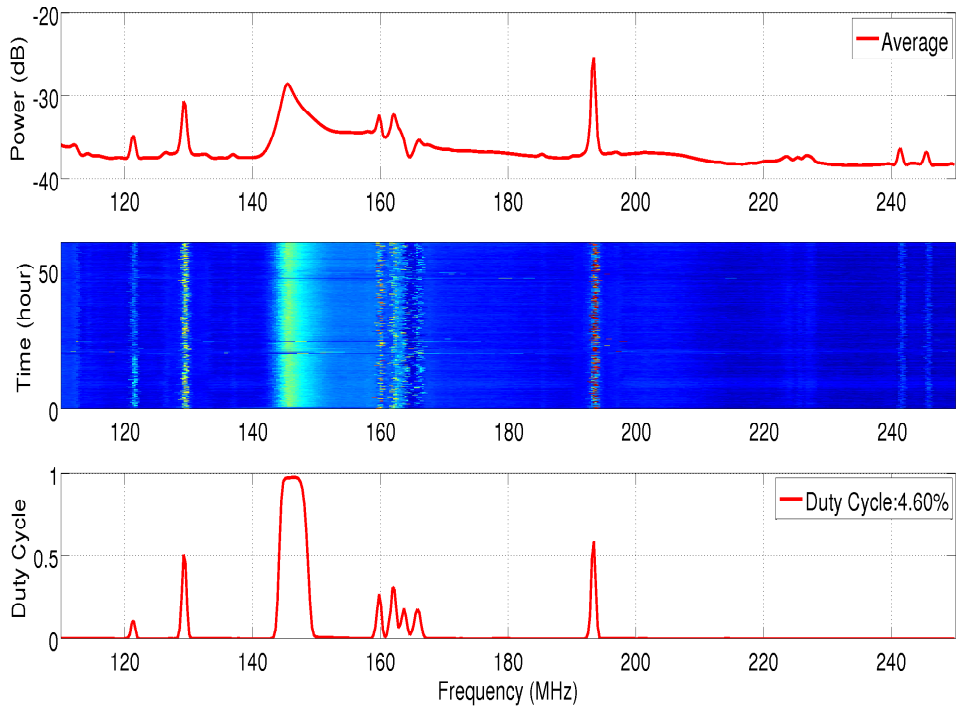

Figure 3.6: Duty cycle and spectrogram of Band 110 to 250 MHz

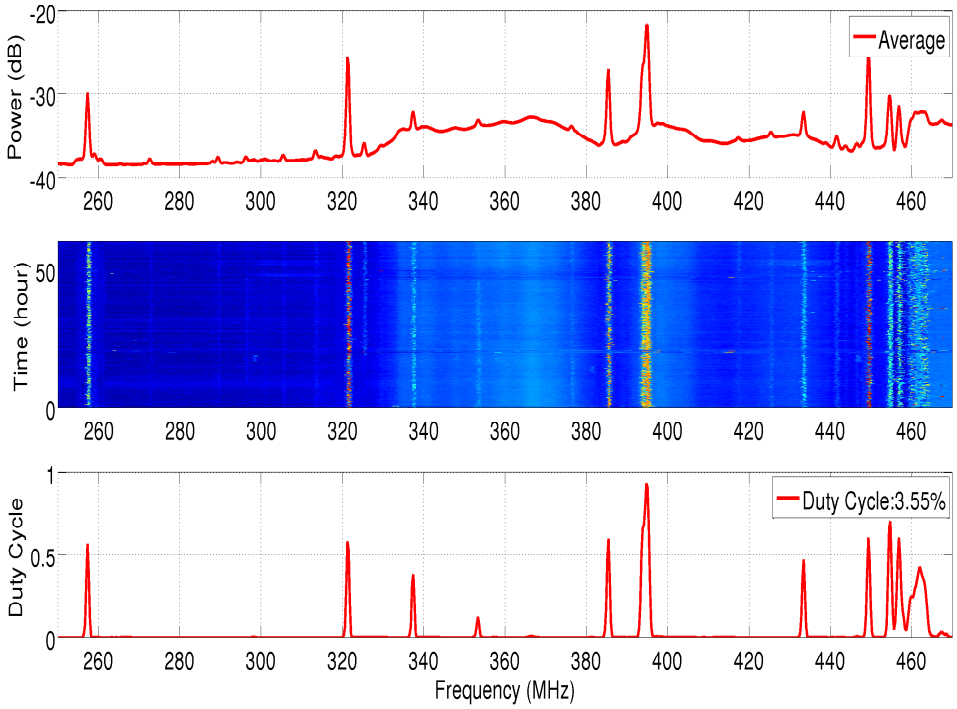

Figure 3.7: Duty cycle and spectrogram of Band 250 to 470 MHz

Table 3.1: Summary of spectrum occupancy at 10dB above ambient noise

Band (MHz)	Services	Average Duty Cycle %	Average Occupied Spectrum (MHz)
50-110	Fixed/Land mobile, Amateur,Aeronautical radio navigation, Radio astronomy,broadcast, FM	3.40	2.040
110-250	Meteorological satellite, Mobile satellite, space research Aeronautical mobile, Land mobile, Maritime mobile Astronomical mobile, Broadcasting, Radio location	4.60	6.44
250-470	Mobile satellite, Space operation. radio astronomy Aeronautical radio navigation, Meteorological aids/Satellite	3.55	7.81
470-600	Broadcasting, TV band IV, Aeronautical radio navigation	28.54	37.102
600-730	TV Broadcast band V, Some PMSE	2.55	3.315
730-860	TV Broadcasting, PMSE, Mobile	15.06	19.578
868-890	IMS 868	14.71	3.236
890-960	GSM 900 band	40.85	28.595
1700-1880	GSM 1800 band	51.57	92.826
2400-2500	ISM 2.4GHz band	18.55	18.550

3.5.1 Results of spectrum occupancy survey for sixteen days

The results of spectrum occupancy survey conducted for a period of sixteen days for bands ranging from 50 MHz to 860 MHz are presented in this section. The RF signals were captured from 23 December, 2015 18:57 to 08 January, 2016 18:56 using USRP1. The ambient noise of each frequency resource block is obtained by replacing the antenna with 50Ω smart antenna terminator and the average of the spectrum captured for 15 hours is used as the noise floor. The threshold is set by adding a constant 10dB margin to the noise floor of each frequency resource block in accordance with ITU recommendation. From the results presented in Table 3.2, the occupancy level of band 50-860 MHz is low.

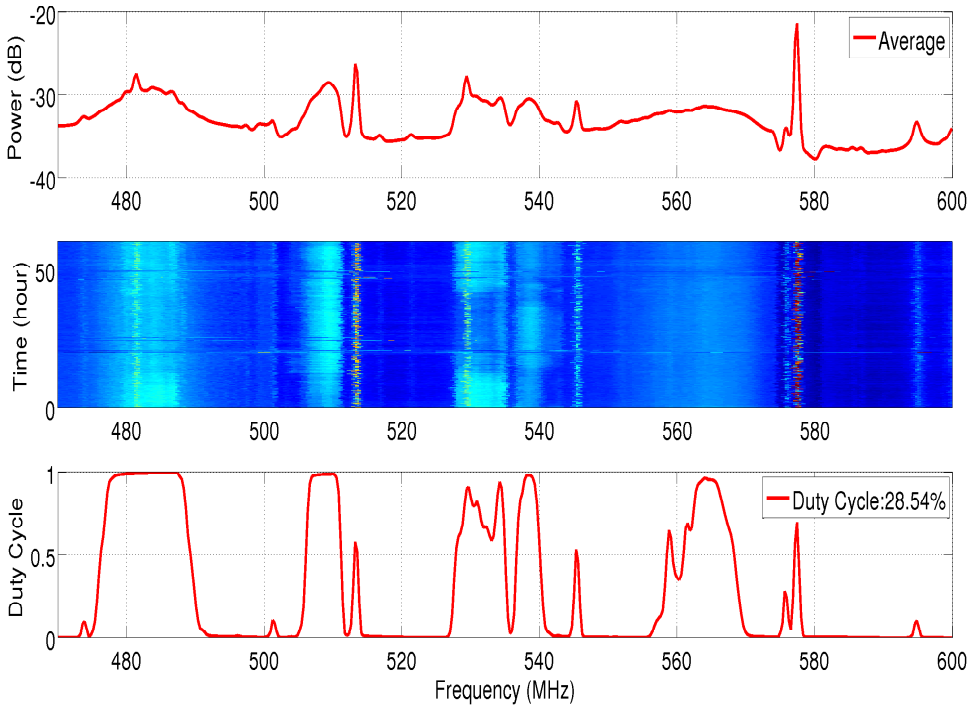

Figure 3.8: Duty cycle and spectrogram of TV Band 470 to 600 MHz

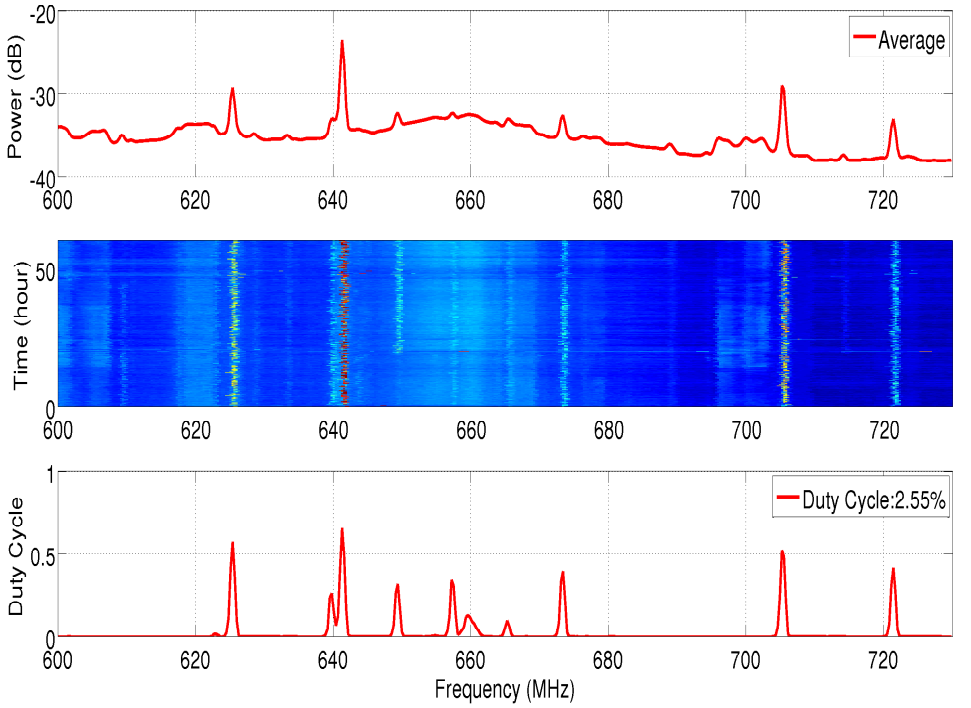

Figure 3.9: Duty cycle and spectrogram of TV Band 600 to 730 MHz

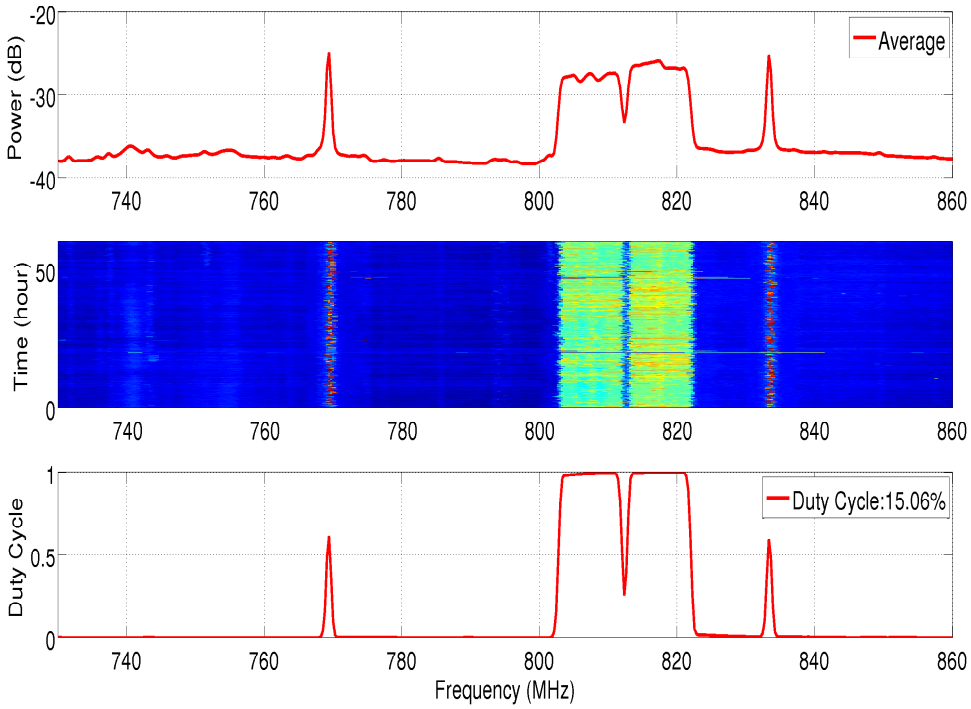

Figure 3.10: Duty cycle and spectrogram TV band 730 to 860MHz

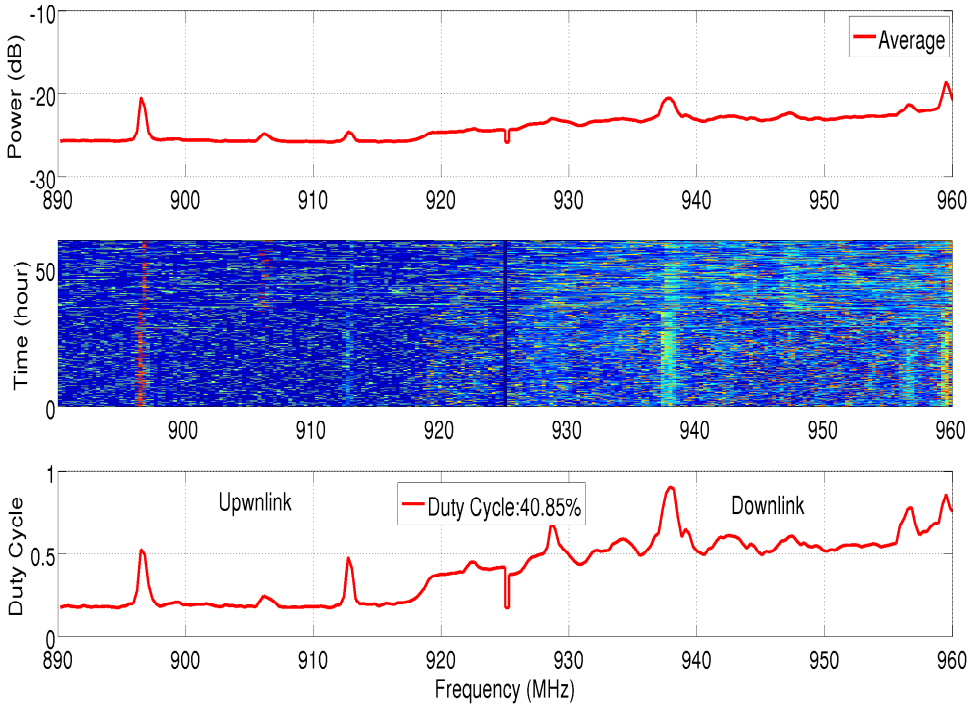

Figure 3.11: Duty cycle and spectrogram of GSM 900 band

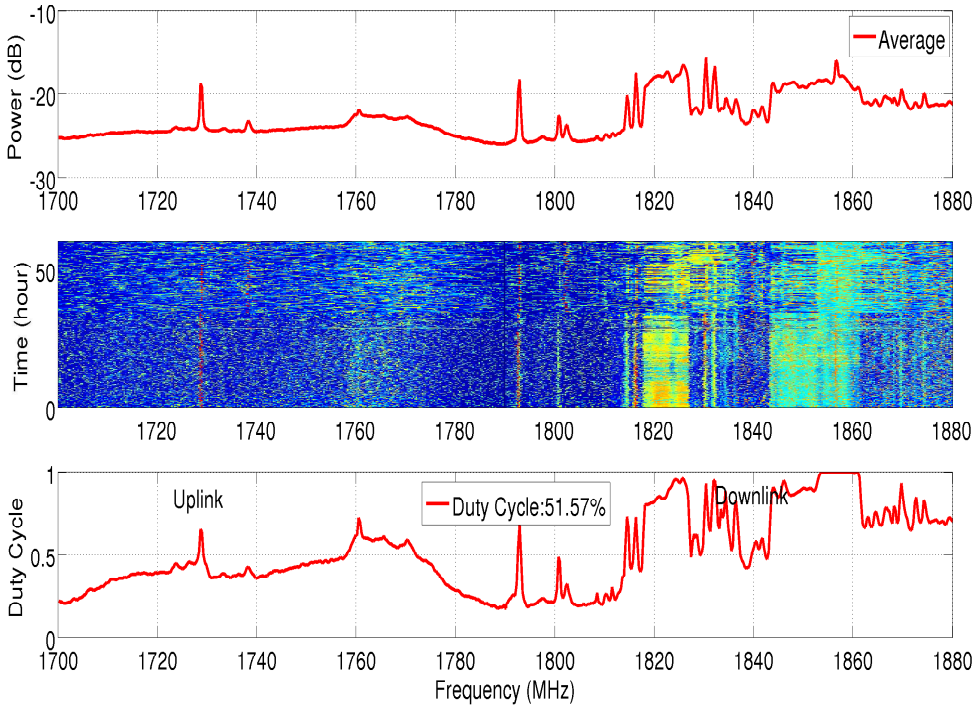

Figure 3.12: Duty cycle and spectrogram of GSM 1800 band

In order to estimate the effect of threshold on the spectrum occupancy, the constant noise margins are set at 5, 7 and 10dB. In both cases the results obtained are highly correlated with the results presented in Table 3.1. The results obtained reveals the significant of the threshold on the spectrum occupancy estimate. Small changes in the threshold has significant impact on the decision outcome of energy detector as shown in Table 3.2. Thus the overall accuracy of energy detector is dependent on the correct choice of the decision threshold.

Comparing the results presented in Figures 3.14, 3.15 and 3.16 which were captured for 16 days from 23 December, 2015 at 18:57 to 08 January, 2016 at 18:56, with the respective bands in Figures 3.5, 3.9 and 3.10 captured for three days from 31 May to 3 June 2015, is obvious that the services allocated to bands ranging from 50 MHz to 860 MHz, seem to have a regular repetitive pattern of transmission; and are confirmed to certain frequencies which is evidenced by the occurrences of the peaks and local peaks of the average power, spectrogram and duty cycles at certain frequencies. The services within this bands seem to operate a routine schedule that make use of certain frequencies continuously while others are completely left idle. This is an important finding as the idle channels can easily be identified for spectrum hole access.

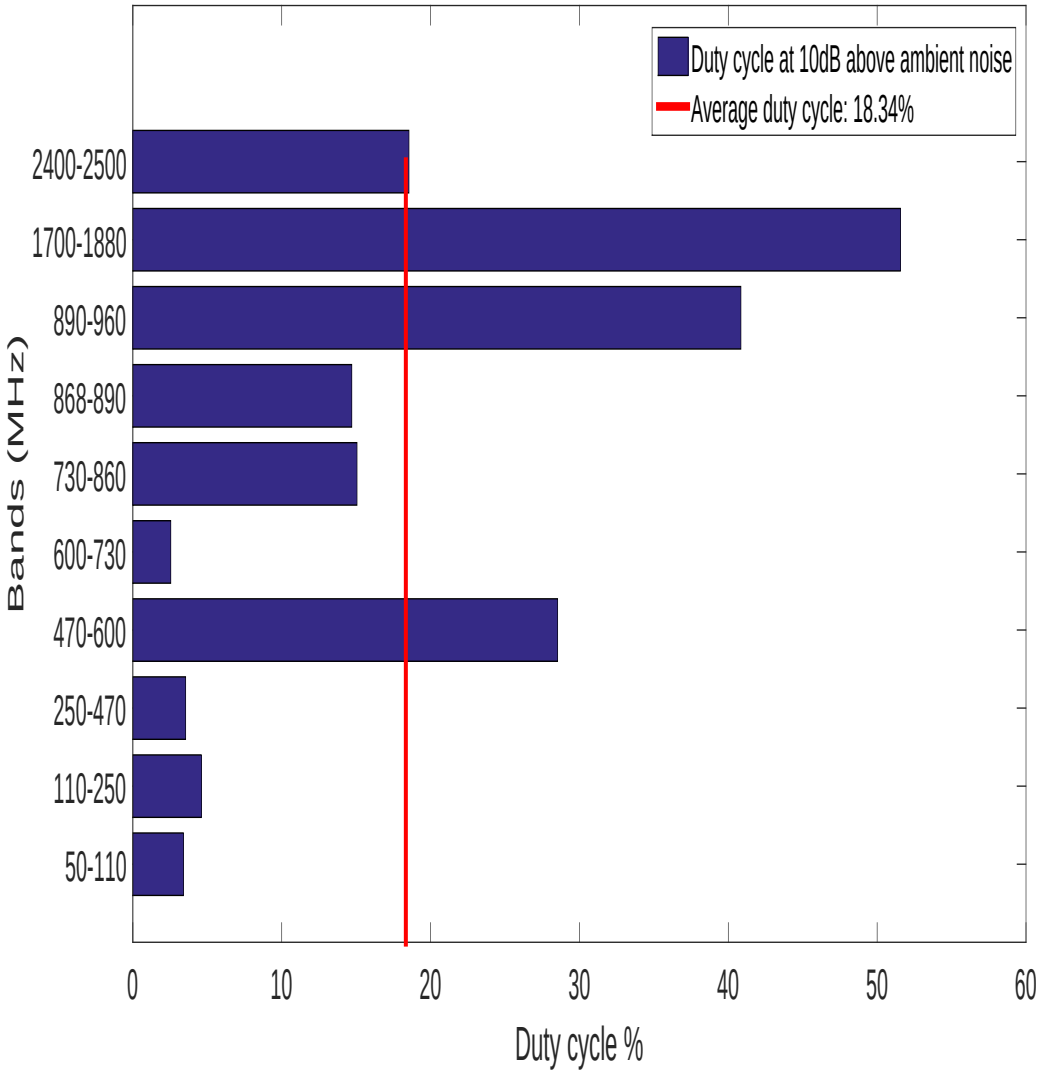

Figure 3.13: Duty cycle at 10 dB above ambient noise

3.5.2 Decision threshold chosen to satisfy a predefined probability of false alarm

Another way of choosing the decision metric is to set the threshold to satisfy a given probability of false alarm, Contreras et al. (2011b). This is achieved by setting the decision threshold in such a way that a certain percentage of the ambient noise power samples are above the threshold. This can be illustrated by Equation (3.13) as follows: In order to obtain the ambient noise power $\varphi(f)$ that satisfied a given probability of false alarm Pf, the RF power of the noise samples w_f of physical resource block (channel) f are ranked in descending order starting from the maximum. The position index I of the noise sample that will satisfy the probability of false alarm is given by Equation (3.11) while the actual noise power is given by Equation (3.12). Where N is the total number of ambient noise power samples and *ceiling* is an operator that returns the least integer

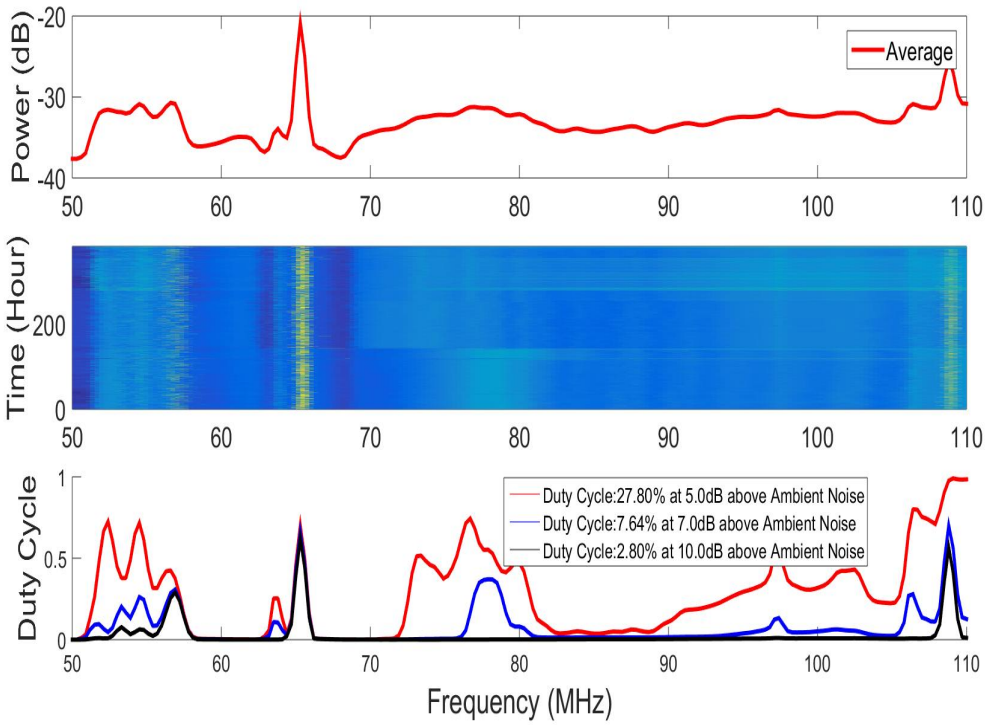

Figure 3.14: Duty cycle and spectrogram of Band 50 to 110 MHz at various noise margins for 16 days

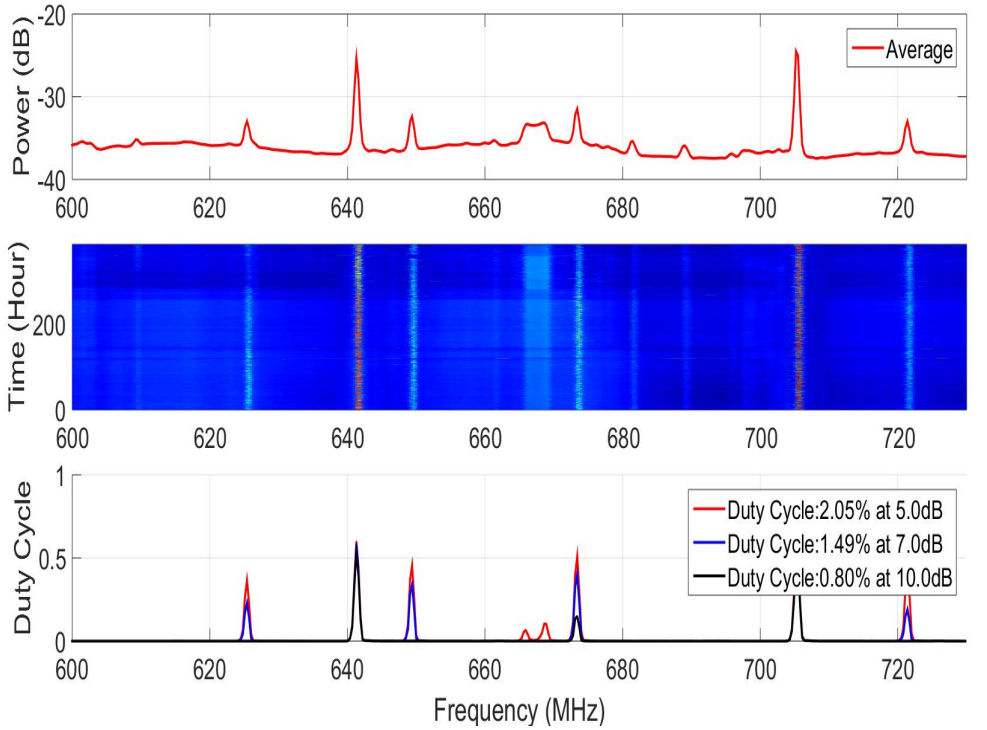

Figure 3.15: Duty cycle and spectrogram of Band 600 to 730 MHz at various noise margins for 16 days

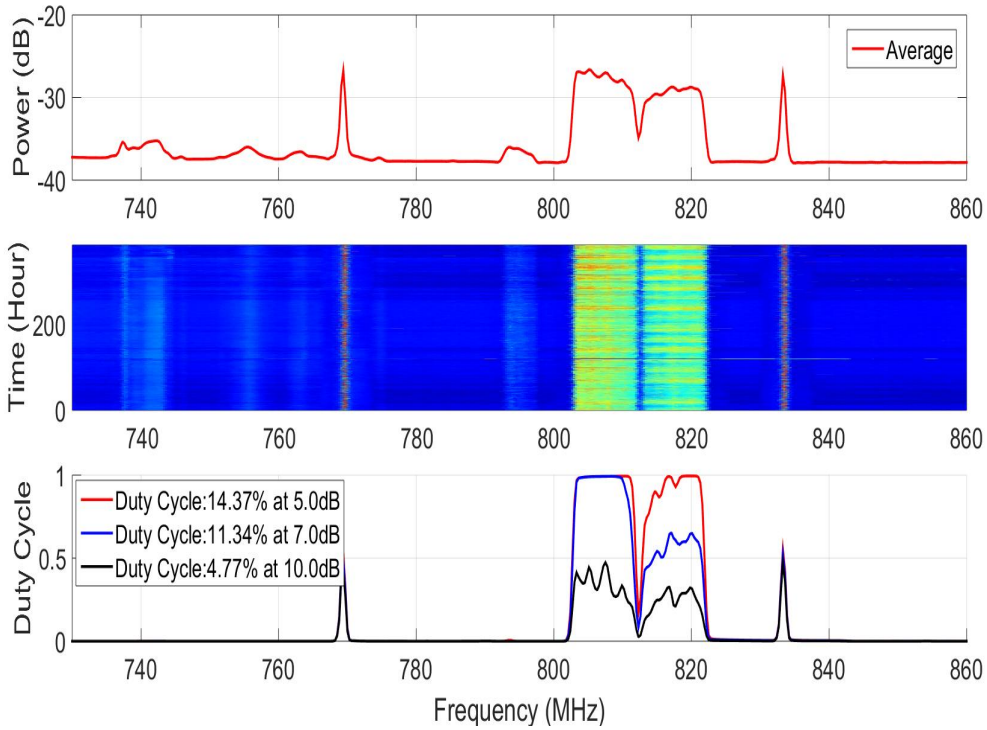

Figure 3.16: Duty cycle and spectrogram of Band 730 to 860 MHz at various noise margins for 16 days

Table 3.2: Summary of sixteen days Spectrum Occupancy

Band (MHz)	Services	Average Duty Cycle (%) at three margins Above Ambient Noise		
		Margin 5 dB	margin 7 dB	margin 10 dB
50-110	Fixed/Land mobile, Amateur, Aeronautical radio navigation, Radio astronomy,broadcast, FM	27.80	7.64	2.80
110-250	Meteorological satellite, Mobile satellite, space research Aeronautical mobile, Land mobile, Maritime mobile,Astronomical mobile, Broadcasting, Radio location	36.11	11.23	4.56
250-470	Mobile satellite, Space operation, radio astronomy, Aeronautical radio navigation, Meteorological aids/Satellite	35.18	21.93	10.15
600-730	TV Broadcast band V, Some PMSE	2.10	1.53	0.84
730-860	TV Broadcasting, PMSE, Mobile	14.41	11.35	4.81

not less than the argument $Pf * N$.

$$I = ceiling(Pf * N) \tag{3.11}$$

$$\varphi(f) = w_f(I) \tag{3.12}$$

$$\top(f) = \varphi(f) + M(f) \tag{3.13}$$

Where $\top(f)$ is the threshold of a physical resource block f, $\varphi(f)$ is the ambient noise power of frequency resource block (bin) f that satisfied the given probability of false

alarm while $M(f)$ is a constant margin whose values depend on the overall desired probability of false alarm for a given resource block f. For this study, $\varphi(f)$ was chosen such that only 1.5% of the ambient noise samples measured for a period of 15 hours are above it. This gives a probability of false alarm of 0.015. The spectrum occupancy estimated using this approach is likely to be more reliable than using the average of the ambient noise as depicted in Section 3.4 and Equation (3.1), especially when the ambient noise variance is high. If the average of the ambient noise is used, and adding the constant noise margin $M(f)$ to obtain the decision threshold, the threshold can fall within the noise range for small $M(f)$ with a consequent increase in false alarm probability. In this second approach with only 1.5% of the noise samples above $\varphi(f)$; when the constant noise margin $M(f)$ is added, the overall probability of false alarm is likely not to exceed 0.015. The extend to which the probability of false alarm will fall below 0.015 depend on the magnitude of $M(f)$. The constant noise margin $M(f)$ was set at 5, 7 and 10 dB, and the results is as shown in Figure 3.17. From Figure 3.17, with $\varphi(f)$ chosen to satisfy 0.015 probability of false alarm at constant noise margins of 5, 7, and 10dB, the average occupancy estimate are 31.52%, 20.86% and 11.70% respectively. These results also depict low level of spectrum utilization just as the ones shown in Tables 3.1 and 3.2 which further unveiled the fact that the current static spectrum allocation policy is an inefficient way of spectrum management. This is because most of the licensed spectrum are underutilised, and by law they are exclusively reserved for use only by the license owners. Secondary users should therefore be allowed to use the free licensed spectrum via CR paradigm for efficient spectrum utilization.

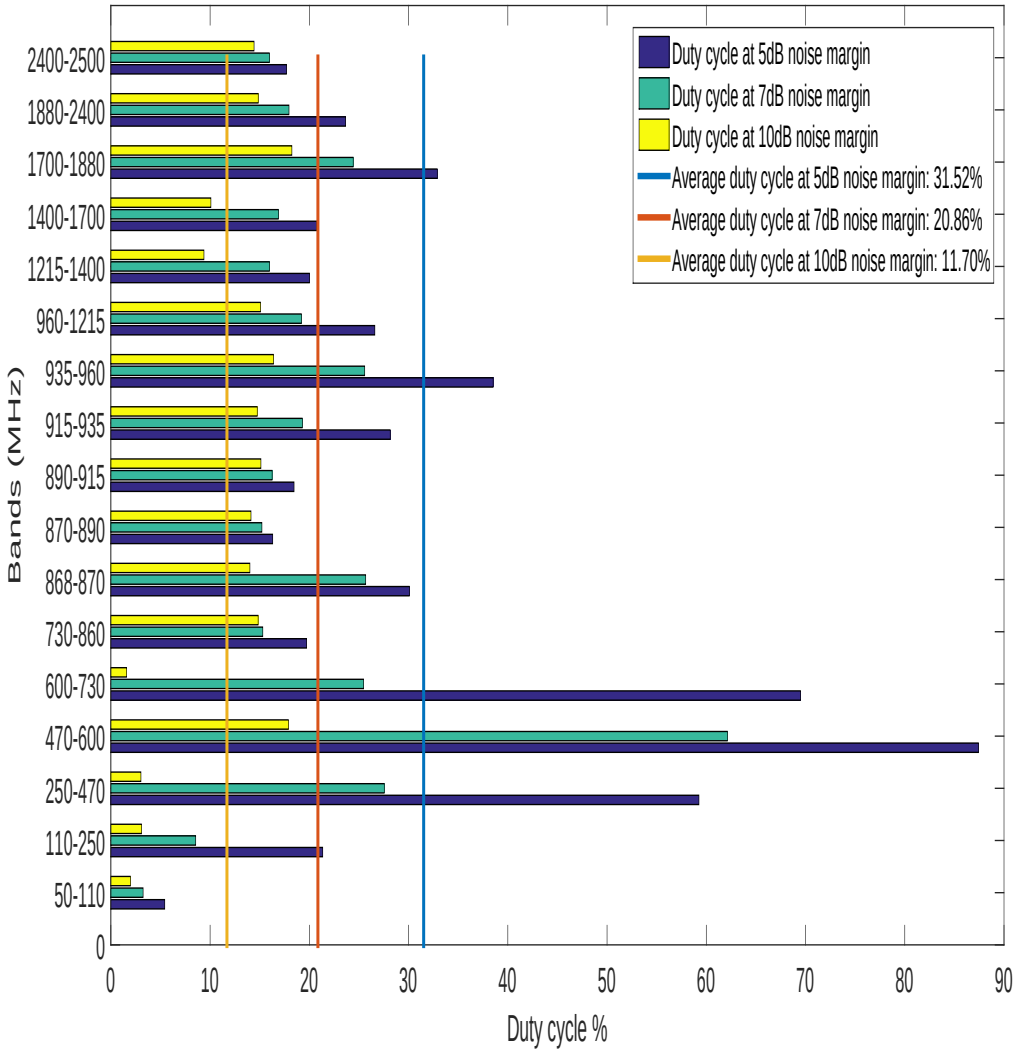

Figure 3.17: Duty cycle at different noise margins and ambient noise of 0.015 probability of false alarm

3.6 Summary of the spectrum occupancy survey

The spectrum occupancy survey presented in this chapter aimed at detecting under-utilized bands that can serve as potential candidate bands to be exploited by cognitive radios (CR). This research demonstrates that bands 50 - 470 MHz are grossly under-utilized with an average occupancy of 3.71%. The duty cycle of band 470 - 860 MHz is 15.70%. GSM 900 and GSM 1800 bands are heavily utilized as compared with other bands examined in this thesis. Based on the findings of the studied location, bands 50 - 860 MHz are potential candidate bands for CR application. As the RF signals are now captured, and the trends of the RF power distribution of the channels, along with their estimated occupancy rate are graphically displayed and studied, the next chapter will present the modelling of RF power and spectrum hole prediction models for efficient

spectrum utilization.

Chapter 4

Spectrum Holes and RF Power Prediction Models

4.1 Introduction

Optimized Artificial Neural Network (ANN) and Support Vector Regression (SVR) models are designed for prediction of real world Radio Frequency (RF) power within the GSM 900, Very High Frequency (VHF) and Ultra High Frequency (UHF) FM and TV bands. The general theory of the prediction models was discussed in Section 2.5. The training and validation dataset are captured using a Universal Software Radio Peripheral (USRP) as explained in Section 3.3. The rest of the chapter is organised as follows: The prediction models specifications and the topologies used in this research are presented in Section 4.2, with Section 4.2.2 and Section 4.2.3 focusing on ANN and SVR models respectively. In order to determine how many past RF power samples should be fed back as part of the inputs of the prediction models, a sensitivity analysis presented in Section 4.3 was conducted. As a means of improving the accuracy and generalization of the prediction model, a combined prediction model consisting of different prediction models was developed. More importantly, a novel benchmark for obtaining the weights of the combined prediction model is proposed as discussed in Section 4.4. The prediction results are presented in Section 4.5 while Section 4.5.3 gives the summary of the prediction results.

4.2 Artificial Neural Network (ANN) and Support Vector Machine (SVM) RF Power Prediction Models

RF power prediction models were developed using ANN and SVM. In order to determine the basic parameters that should serve as the input of the prediction models, the results of some spectrum occupancy experiments conducted by other researchers using energy detector were considered and reviewed. One interesting observation reveals that there are seasonal (e.g. Christmas, new year, holiday period, etc), weekly (e.g. week-

days and weekends) and daily trends associated with spectrum occupancy statistics, Bacchus et al. (2010). A comprehensive wide band spectrum occupancy campaign was conducted Qaraqe et al. (2009), where spectrum measurement were taken in four different places simultaneously for frequency band 700MHz - 3GHz reveals that the spectrum occupancy and utilization varies with time, frequency and spatial domain (geographical location). This work shows that spectrum occupancy is strongly dependent on the geographical location. Another study conducted with reference to TV band shows that the spectrum occupancy within the same geographical location conducted at the same time is not only dependent on the location but it also depends on whether the measurements were taken indoor or outdoor, Contreras et al. (2011a). This study shows that there is more spectrum occupancy outdoor than indoor. From the same study, it was found that there is a clear difference between the day time and night time spectrum occupancy estimates within the investigated TV band. In addition to the aforementioned, spectrum occupancy also depend on the altitude or the height above the sea level where the measurements are taken. Thus even if the measurements were taken at the same geographical location within the same time frame and frequencies, but at different heights, the spectrum occupancy will not be the same. In order to have a robust prediction models with reduced dimensionality of input vector, the inputs of the proposed RF power prediction model consists of time, frequency, longitude, latitude, height and some past recent (feedback) RF power samples as detailed in Sections 4.2.2 and 4.2.3. Sensitivity analysis discussed in Section 4.3 was used for determination of the optimum number of past recent RF power samples to be used as part of the input of the ANN and SVM for prediction of current RF power.

4.2.1 Performance index

Once the types of prediction models to be used are chosen, and the corresponding inputs and outputs of the models are identified, there is need for correct choice of performance metric that will be used to estimate how well the prediction models are able to solve, and capture the trends associated with the problem. There are many indexes or empirical risk functions (error measurements) that can be used to evaluate the performance of a prediction model, each with it merits and demerits. The correct choice of performance index plays a crucial role in selecting the most accurate and generalised model. These performance metrics include:

- Mean Absolute Error (MAE)

  This error index has the same unit as the data. As the errors are not squared, it is not sensitive to outliers as compared to mean square error (MSE). It is similar to root mean square error (RMSE) but slightly smaller in magnitude. This metric is used to quantify the forecasting error in terms of the measured unit, Equation (4.1).

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|P_t - \hat{P}_t| \tag{4.1}$$

Where N is the number of observations (RF power samples), t is a time index in the range $t = 1, \cdots, N$ while P_t and $\hat{P}_t$ are the actual and predicted RF power at time t.

- Mean Absolute Percentage Error (MAPE)

  This error metric is expressed in percentage hence it is dimensionless and is independent of the scale or unit of the data. It can easily be understood even by non expert and can be used to compare models with different output data unit (scale). It can also be used where the scale (unit) of the time series data changes with time but it can not be used where the time series data can take zero. Also MAPE does not make sense when the time series data is negative. Since MAPE is relative to the actual value, the result can be miss leading (i.e. very large) if the actual data is very small or near zero, Equation (4.2).

$$MAPE = \frac{100}{N} \sum_{t=1}^{N} \left| \frac{P_t - \hat{P}_t}{P_t} \right| \tag{4.2}$$

- Mean Square Error (MSE)

  MSE is highly sensitive to outliers as the error is squared, Equation (4.3). This metric is expressed as the square of the actual unit which gives disproportionate weight to a very large error while errors with magnitude less than one are made much more smaller. If the true risk of error is proportional to the square of the error, then MSE is a good choice for model evaluation. For this study, since the measured RF power is in Decibel (dB) which can be any real number including zero and negative numbers, MAPE cannot be used. In order to have prediction models with good performance in the events of outliers, the MSE is used to evaluate the performance of the prediction models implemented in this research.

$$MSE = \frac{1}{N} \sum_{t=1}^{N} (P_t - \hat{P}_t)^2 \tag{4.3}$$

- Root Mean Square Error (RMSE)

  RMSE is the square root of MSE, Equation (4.4) and is expressed in the same unit as the measured variable. It is less sensitive to errors due to outlier data points as compare with MSE. The RMSE is a good metric for model selection where the risk resulting from the error is proportional to the error. This has been used for comparing time series forecasting models, Almaraashi et al. (2010).

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (P_t - \hat{P}_t)^2} \tag{4.4}$$

4.2.2 Artificial Neural Network

The Artificial Neural Network (ANN) is trained using differential evolutionary and particle swarm algorithms variants after which the weights are further fine tuned using backpropagation algorithm (BPA). The training objective function is the minimization of the mean square error (MSE) i.e. the synaptic weights and biases were updated every epoch to minimize the MSE. A supervised batch training method was used with 60% of the data used for training the ANN, 20% for validation and 20% for testing the trained ANN. A large proportion of the data (60%) was used for training in order to build a robust network capable of capturing all possible scenarios or trends associated with the RF power traffic of the studied location. The inputs of the ANN consist of four past recent RF power samples, channel identifier, and time domain data of varying rates of change i.e. second, minute, hour, week day (1 to 7), date day (1 to at most 31), week in a month (1 to 5), and month (1-12) while the output gives the RF power in Decibels (dB). Each input of the time domain, enables the ANN to keep track with the trend of RF power variation as a function of that particular input. The current RF power is modelled as a nonlinear function of recent past RF power samples, location and current time, thus forming a nonlinear time series model. Splitting the time domain input data into time segments of various rate of change increases the input dimension and training time, but it was experimentally verified to result in a model that is more generalized than when trained with the time domain data fed as a single stream through one input. The number of past samples to be used (in this study 4) for reliable prediction and efficient memory management was obtained experimentally as detailed in Section 4.3. The actual past RF power samples fed at the input of the ANN, coupled with the long time training information captured via the time domain inputs, results in a robust ANN model that adapt well to the present trend of RF power variation. In this study, three ANN models were designed. The first model is shown in Figure 4.1; it consists of only one output neuron and is dedicated for RF power prediction of only one channel which implies that each channel will have its own dedicated ANN. To circumvent this problem, two other models are designed for RF power prediction in multiple channels. The second model depicted in Figure 4.2 is used for prediction of RF power in many channels (for this study is 20 channels) but one at a time. This model has only one output neuron, but in addition to the time, and past RF power samples inputs, it has another input representing the channels. The output neurons of the third (parallel) model is equal to the number of channels to be considered, Figure 4.3. The parallel model is used for simultaneous prediction of RF power in multiple channels given the current time instant and past RF power samples as inputs. For the parallel model, if 4 recent past samples of each of the channels were used as distinct feedback inputs, there will be a total of $4 \times N$ feedback inputs; where N is the number of channels, Figure 4.3; and the training will be computationally expensive. These large feedback inputs ware reduced to 4 by using their average.

The data used in this study were obtained by capturing real world RF signals as discussed in Section 3.3. In all the models, no RF power related parameters such as

signal to noise ratio (SNR), bandwidth, and modulation or signal type, etc, are used as the input of the ANN. Thus making the models robust for cognitive radio application where the CR has no prior knowledge of these RF power related parameters. Four ANN topologies were considered: feed forward (FF), feed forward with output feedback (FFB), cascaded feed forward (CFF) and layered recurrent (LR) ANN.

The accuracy and level of generalization of an ANN depends largely on the initial weights and biases, learning rate, activation functions, momentum constant, training data and also the network topology. In this research, since the backpropagation algorithm (BPA) is used as a local searcher, the learning rate and the momentum were kept low at constant values of 0.01 and 0.008 respectively while the initial weights and biases were evolved using differential evolutionary and particle swarm algorithm variants. The first generation initial weights and biases were randomly generated and constrained within the decision space of -2 to 2 using a toroidal bound as described in Section 2.6.3. The decision space (-2 to 2) is obtained experimentally after a series of manually turning. After 500 generations, the ANN weights and biases were initialized using the global best i.e. the most fittest solution (candidate with the least MSE obtain using test data) and then train further using backpropagation algorithm (BPA) to fine tune the weights toward the global optimum as detailed in the training Section 4.3.1. Thus producing the final optimized ANN model. The activation function $f(z)$ used for the hidden layers is as shown in Equation (4.6) while the output activation function is depicted by Equation (4.7).

$$z = \sum_{i=1}^{q} w_i x_i + b \tag{4.5}$$

$$f(z) = \frac{1 - e^{-z}}{1 + e^{-z}} \tag{4.6}$$

$$f(z) = K \cdot z \tag{4.7}$$

Where z is the activation, x is a multivariate q dimensional input vector, w are the weights, b is the bias, $f(z)$ is the activation function and K is the gradient, as discussed in Section 2.5.1.

4.2.3 Support vector machine

The general theory of SVM regression model has been presented in Section 2.5.2. In this study five SVM regression models are designed using five kernel functions i.e. the Radial Basis Function (RBF), Gaussian Radial Basis Function (GRBF), Exponential Radial Basis Function (ERBF) kernel, Linear Kernel (LK) and Polynomial kernel (PK) given by Equations (4.8), (4.9), (4.10), (4.11) and (4.12) respectively, Haykin (2008).

- Radial basis function kernel:

$$K(x, x_i) = e^{(-\gamma \|x - x_i\|^2)} \tag{4.8}$$

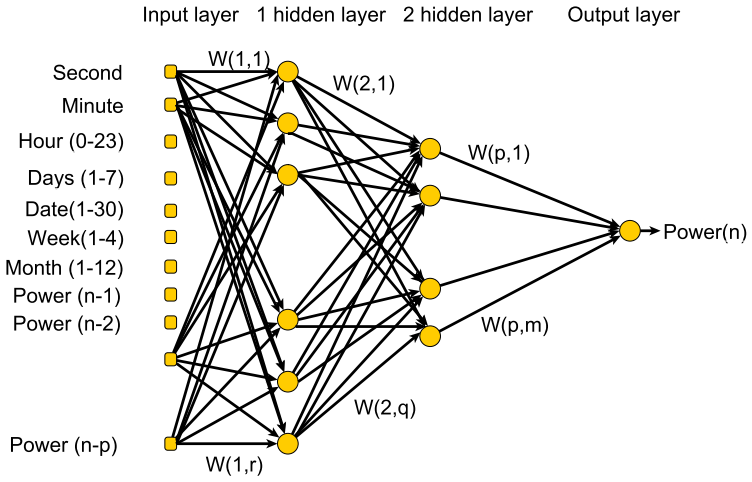

Figure 4.1: Dedicated ANN model for one channel

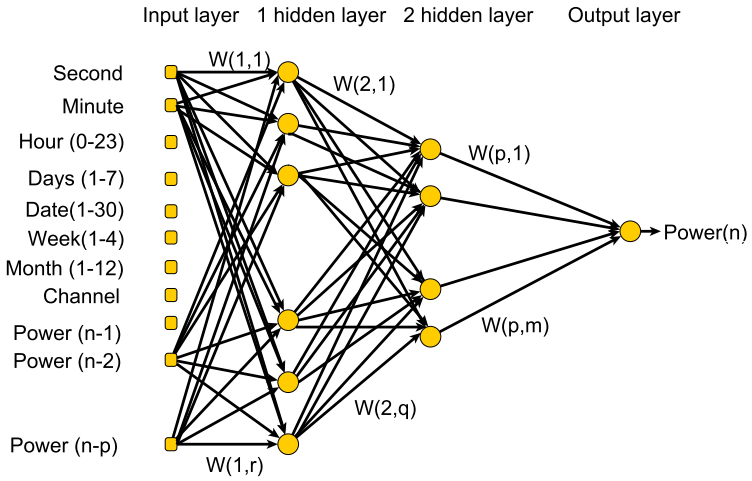

Figure 4.2: Multiple channels, single output ANN model

Where n is a time index, $Power(n-1), Power(n-2), \cdots, Power(n-p)$ are the past p RF power samples while $Power(n)$ is the current predicted RF power.

- Gaussian kernel:

$$K(x, x_i) = e^{(-\frac{\|x-x_i\|^2}{2\sigma^2})} \tag{4.9}$$

- Exponential kernel:

$$K(x, x_i) = e^{(-\frac{\|x-x_i\|}{2\sigma^2})} \tag{4.10}$$

- Linear kernel:

$$K(x, x_i) = x^T x_i + c \tag{4.11}$$

- Polynomial kernel:

$$K(x, x_i) = (\alpha x^T x_i + c)^d \tag{4.12}$$

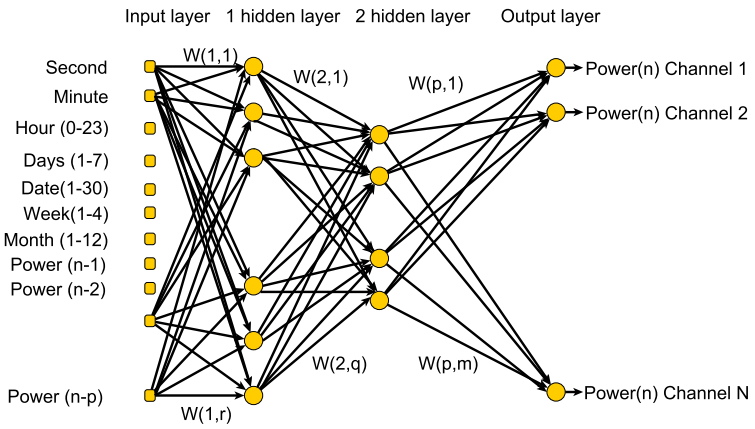

Figure 4.3: Multiple channels, parallel outputs ANN model

Where n is a time index, $Power(n-1), Power(n-2), \cdots, Power(n-p)$ are the past p RF power samples while $Power(n)$ is the current predicted RF power.

The empirical risk $R_{emp}(w)$ to be minimised in SVM regression model is given by:

$$R_{emp}(w) = \frac{1}{n}\sum_{i=1}^{n} L_{\varepsilon}(y_i, f(x_i, w)) \tag{4.13}$$

Support vector regression model is formulated as the minimization of the following objective functions, Haykin (2008):

$$\text{minimise} \quad \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i{}^*) \tag{4.14}$$

$$\text{subject to} \quad \begin{cases} y_i - f(x_i, w) - b \leqslant \varepsilon + \xi_i{}^* \\ f(x_i, w) + b - y_i \leqslant \varepsilon + \xi_i \\ \xi_i, \xi_i{}^* \geqslant 0, \quad i = 1, \cdots, n \end{cases} \tag{4.15}$$

Where all the parameters retain their meaning as explained in Section 2.5.2. The novel approach for obtaining the metatparameters of the SVM regression model is detailled in Section 2.7.

Two SVR models were implemented for each kernel, one of the models depicted in Figure 4.4 is dedicated for prediction of RF power of only one channel or resource block which implies that each channel will have its own model; the second model shown in Figure 4.5 has an additional channel input which is designed for prediction of RF power in many channels but one at a time. When the SVR was trained with the time domain input data splitted into time segments of varying rate of change (second, minute, hour, date day, day, week, month) just as the ANN as shown in Figures 4.1 and 4.2, the evolved SVR model performance was poor (fail to generalize), this may due to large evolved support vectors as a results of increased input vector dimension. When the time domain input data was fed to the SVR as a single stream of data, the SVR model performance increases significantly. Hence the SVR models implemented has only one

input for time while the time inputs of the ANN were split as shown in Figures 4.1 to 4.3.

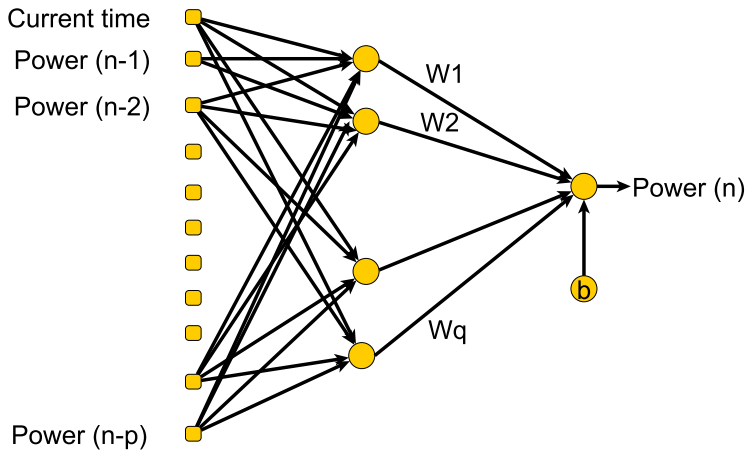

Figure 4.4: Dedicated SVM model for one channel

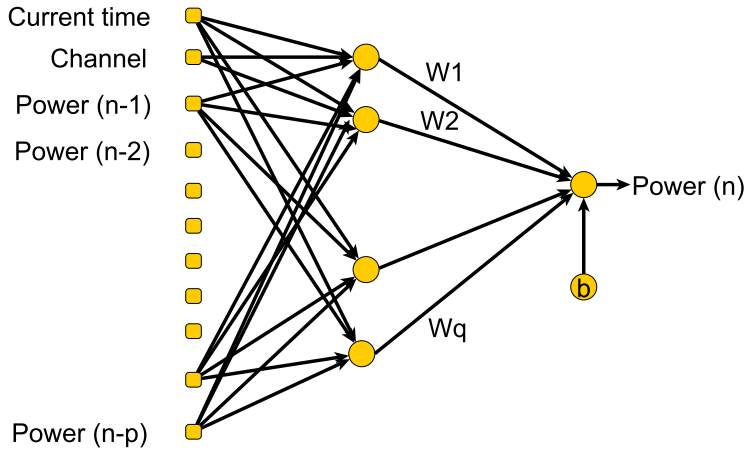

Figure 4.5: Multiple channels, single output SVM model

Where n is a time index, $Power(n-1), Power(n-2), \cdots, Power(n-p)$ are the past p RF power samples while $Power(n)$ is the current predicted RF power.

4.3 Delayed Inputs Sensitivity Analysis

In order to examine how many numbers of past RF power samples are needed as feedback inputs for reliable prediction, and to have a model with reduced dimensionality of input vector, sensitivity analysis was conducted with reference to the inputs of the ANN and SVR models. One way of evaluating the importance (significance) of an input in ANN is to measure the Change Of MSE (COM) when the input is deleted or added to the neural network or SVM, Sung (1998). In this study, the COM method is adopted with the time domain inputs unaltered, and the actual past RF power are added to the

input one after the other starting from the most recent one. The sensitivity analysis of the ANN input was conducted as follows: the ANN is first trained with i delay inputs (past RF power samples) and the MSE MSE_i is evaluated. The network is retrained with $i+1$ delayed inputs, and the new MSE MSE_{i+1} is obtained. The change in the MSE, $\delta_{mse} = MSE_{i+1} - MSE_i$ is computed as a means of evaluating the importance of the $i+1$ delay input, for $i = [0 \cdots p]$, where p is the total number of past RF power samples used. Note, δ_{mse} is not computed relative to the MSE obtained when all the p delayed inputs are used as in normal COM method, due to the fact that the required number of delay inputs p for achieving a given performance index is not known at the start of the experiment; in this case p is obtained by setting a constrain on δ_{mse}. The importance of the inputs are ranked base on the one whose addition causes the largest decrease in MSE as the most important since they are most relevant to the construction of a network with the smallest MSE. In order to justify the importance or ranking of the inputs statistically, for every i^{th} inputs delay, the ANN is trained 20 times, each time with a randomly generated initial weights and biases, the average of the 20 MSE_i is used. The ranking using the normalized values of change in average MSE δ_{mse} as delayed inputs were added is as shown in Figure 4.7. The graph of the average MSE against number of delayed inputs is as depicted in Figure 4.6. From Figure 4.6 and Figure 4.7, it is obvious that when the number of delay inputs is greater than 7, the change in MSE δ_{mse}, is very small and insignificant. Instead of using 7 past RF power samples as part of the input of the ANN; to further reduce the dimension of the input vector, the percentage of positive change (decrease) in MSE was computed, the result shows that the first four most recent RF power samples account for 94.68% of the total decrease in MSE resulting from all the past RF power samples used. Thus in this study, 4 past recent RF power samples are used as part of the ANN inputs for current RF power prediction. To avoid over-fitting, the results depicted in Figure 4.6 and Figure 4.7 were obtained using the test data, and the fact that the MSE decreases with increase in delay inputs shows that there is a high probability that the ANN may not suffer from over-fitting up to the 14^{th} delay input considered in this study.

When the sensitivity analysis was carried out using SVR model for four kernel as depicted in Figure 4.8, it is interesting to observe that after the 4^{th} delay input, the positive change (decrease) in the average MSE for the four kernels investigated is approximately zero up to the 10^{th} and 14^{th} delay inputs for RBF kernel, and the other three kernels (Gaussian RBF, exponential RBF, linear kernel) respectively. The MSE of the RBF kernel does not maintain a regular pattern after the 10^{th} delayed input while the other three remains fairly constant from the 4^{th} up to the 14^{th} delayed input. Hence a fair choice for the number of past RF power samples to be fed as input to the SVR model for these four kernels should range from four to ten. To minimise the chance of having an over-fit model and to have a model with reduced input dimension, four most recent past RF power samples are used as part of the input of the SVR model. From the results of the sensitivity analysis conducted using both the ANN and SVR model, it seem that the first four most recent past RF power samples bears strong correlation to the models prediction accuracy. For this study, in addition to the time domain and

channel inputs, four most recent past RF power samples are used as part of the input to the ANN and SVR models.

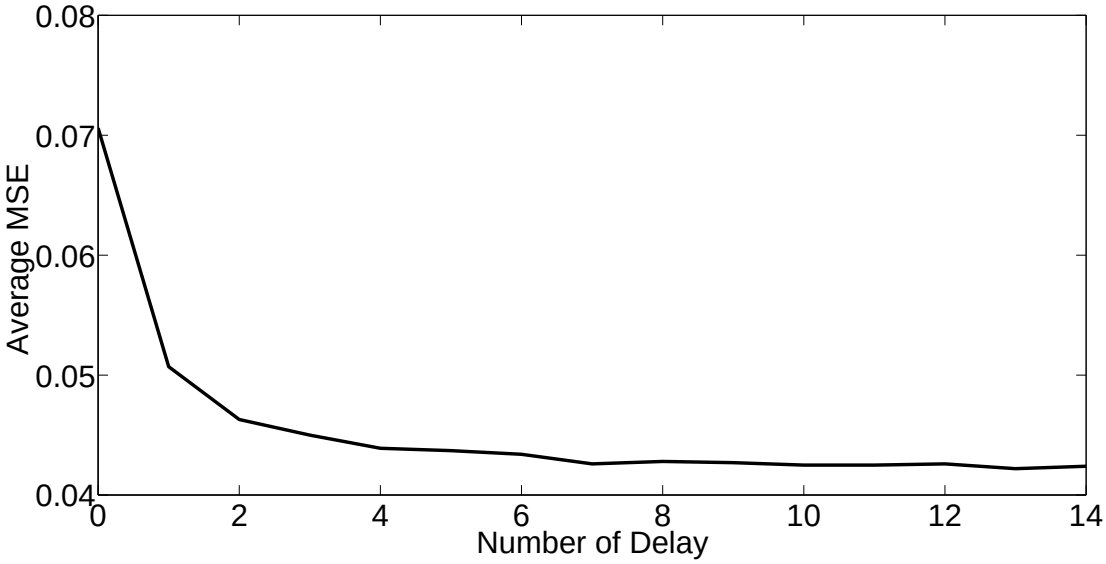

Figure 4.6: ANN sensitive analysis curve as a function of past RF power samples used as inputs

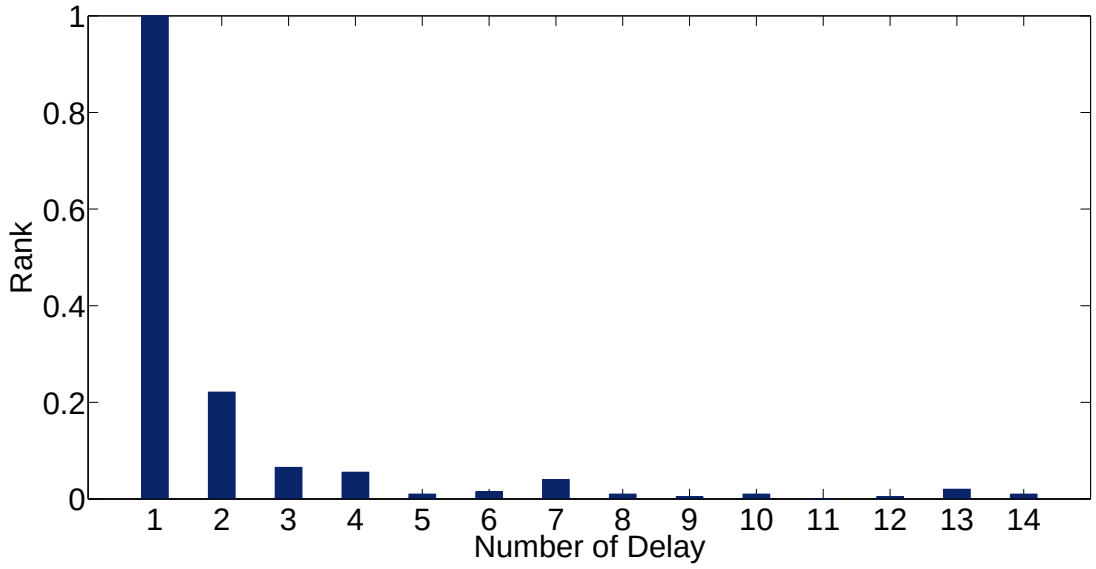

Figure 4.7: ANN past RF power sensitivity ranking

4.3.1 Training of ANN and SVM

One of the desirable features of back propagation algorithms (BPA) is it simplicity, but it often converges slowly and lacks optimality as it can easily be trapped in a local optimum leading to premature convergence. Many approaches has been adopted to solve the problem of premature convergent associated with BPA such as the introduction of

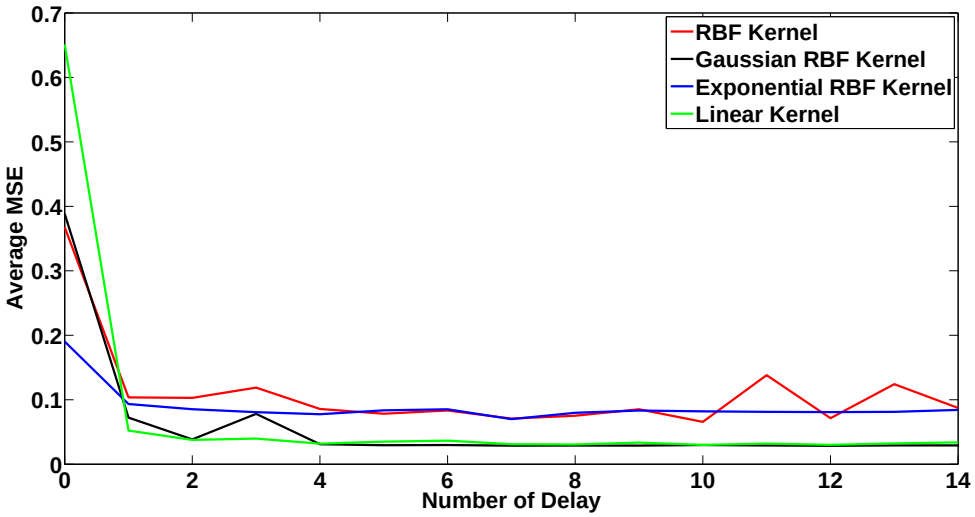

Figure 4.8: SVR sensitive analysis curve as a function of past RF power samples used as inputs.

a momentum constant, varying of the learning rate and retraining of the network with new initial weights. To circumvent the problem of premature convergence, and to have a robust ANN that is well generalized, the global search advantages of population based randomization optimization algorithms and the local search capability of single solution BPA were combined to evolve the weights and biases of the ANN. The initial weights were evolved using differential evolution (DE) and particle swarm optimization (PSO) variants detailed in Section 2.6 and finally fine tuned using BPA towards the global optimum. The objective function in this study is the minimization of the MSE computed using the test data. For the DE, after every generation G, and for each candidate solution i, the offspring $U_{i,G}$ and its parent $X_{i,G}$ are used to set the weights and biases of the ANN and the MSE of the ANN models are computed. The use of the test data for computation of the fitness function (MSE) does not only result in a more accurate network but also a more robust and generalized ANN model. A greedy selection schemes is used in which if the MSE of the offspring is less than that of its parent, the offspring will replace its parent in the next generation otherwise the parent will be allowed to seed the next generation. While for the PSO, if a particle outperform its local best, the local best will be replaced. At the end of the generation, the most fittest individual (global best) i.e. the candidate with the least MSE among the final evolved solutions is used to initialize the weights and biases of the ANN which is further trained using back propagation algorithm in order to produce the final optimized ANN model. The fine tuning of the ANN weights using BPA was constrained within a maximum of 50 epoch and 6 validation fails, i.e. the training stop if any of these constrain thresholds is satisfied or the MSE is within the acceptable set range. The training or estimation data were the only known data sources used in training the ANN. The test dataset was unknown to the network i.e. they are not used in training the network rather are used in testing the trained ANN as a measure of the generalization performance of the ANN

model.

In contrast to the training of the ANN using BPA, the training of SVM is optimal with the optimality rooted in convex optimization. This desired feature of SVM is obtained at the cost of increased computational complexity. The fact that the training of SVM is optimal does not imply that the evolved machine will be well generalized or have a good performance. The optimality here is based on the chosen meta-parameters (i.e. C parameter, ε and the kernel parameter γ), the type of kernel function used and the training data. The SVM is trained using a hybrid ad-hoc nested algorithm whose outer layer consists of different variants of population based and compact metaheuristic optimization algorithms for sampling the SVM meta-parameters (C parameter, ε and the kernel parameter γ), while the inner layer is a convex optimization algorithm for evolving the weights and bias of the SVM. The decision (search) space of the meta-parameters are obtained using the improved approach presented in Section 2.7. At each generation, the meta-parameters are set using each candidate solution, and the corresponding weights and biases are computed using convex optimization algorithms. In order to estimate how the SVM will generalize to an independent dataset (test data), two fold cross validation commonly known as holdout method is used. This has the advantage of having both large training and validation datasets, and each data point is used for both training and validation on each fold. The training data is randomly divided into two sets e.g. A and B of equal size. The SVM was trained on A and test on B, after which it is trained on B and tested on A, the average of the MSE for the two test was used as the fitness function for the given sets of meta-parameters. At the end of the generation, the SVM is reconstructed using the most fittest meta-parameters.

4.3.2 Tool developed for designing prediction models and spectrum occupancy survey

The tool developed for designing the prediction models and for estimation of the spectrum occupancy of the channels is presented in Figure 4.9. The analysis tool is developed using matlab.

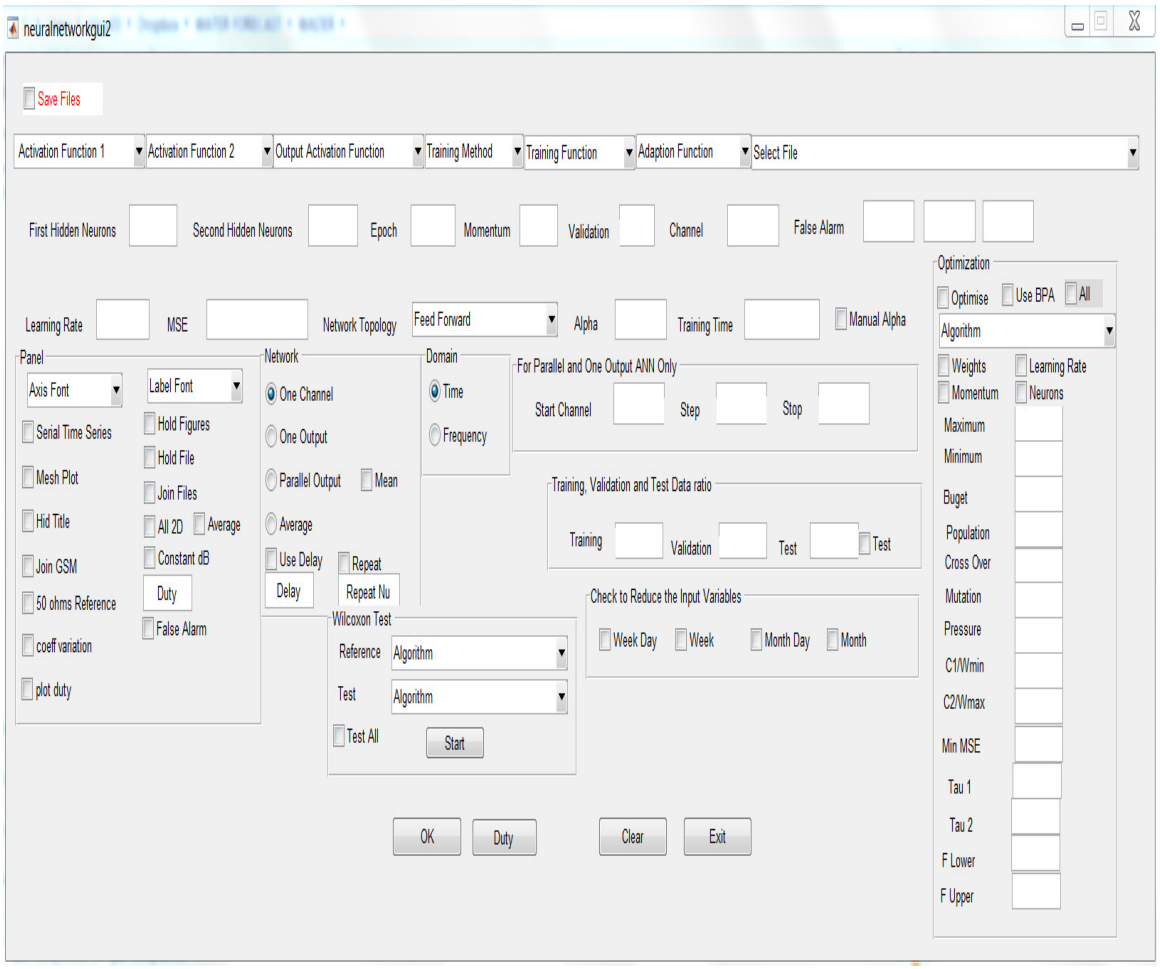

Figure 4.9: Matlab GUI for implementing the prediction models, and for estimating the spectrum utilization of the bands

4.4 Combine prediction model

The accuracy and generalization of a prediction model can be improved significantly by proper combination of two or more models to form a single prediction model Yin-shan et al. (2011). The empirical risk function to be minimized is the MSE, Equation (4.3). There are many methods adopted for selection of the weights of a combine forecasting model. Some of the commonly used methods for static weights selection are: equal weight average method, inverse MSE method and inverse ranking method. In the simple average method, also referred to as equal weights average method, all the output of the models are weighted equally. In this case, the output of each model is weighted with the reciprocal of the number of models i.e. $\frac{1}{m}$, this is the same as taken the average of the outputs of the individual prediction models that formed the combine model, Equation (4.21). The weighted outputs are sum together to give the final output of the combined model. In the inverse MSE method, the weight of each model is inversely proportional to its MSE as shown in Equation (4.22). In inverse ranking method, the models are ranked based on their MSE starting from the one with the least error, i.e. the model with the least error or least empirical risk function is rank first with a

numerical rank of 1, the second best model is assigned rank 2 and so forth. The weights are inversely proportional to the rank of the models as depicted in Equation (4.23). Another method often used in obtaining weights of combine prediction model is the least square method where the weights are the least square estimate of Equation (4.18). The performance of the combine model depends on the evolved weights. Optimal selection of the weights for a given combine prediction model is challenging, thus in this study different variants of differential evolution (DE) algorithms are used to evolve the weights. Detail explanation of how the individual models are aggregated to form one combine model is as follows: Assuming P_t is the actual RF power at time t where $t = 1, \cdots, n$, and $\hat{P}_{i,t}$ be the predicted RF power of model i at time t where $i = 1, \cdots, m$; n is time index and m is the number of prediction models combine to form one model. if $e_{i,t}$ is the prediction error of model i and E_t is the combine model error at time t, then the combine prediction model optimization problem can be formulated as follow, Yin-shan et al. (2011) :

$$e_{i,t} = P_t - \hat{P}_{i,t} \tag{4.16}$$

$$Q_t = \sum_{i=1}^{m} \phi_i \hat{P}_{i,t} \tag{4.17}$$

$$E_t = P_t - Q_t = \sum_{i=1}^{m} \phi_i e_{i,t} \tag{4.18}$$

$$\text{minimise} \quad \sum_{i=1}^{n} {E_t}^2 \tag{4.19}$$

$$\text{subject to} \quad \sum_{i=1}^{m} \phi_i = 1, \quad \phi_i \geqslant 0 \tag{4.20}$$

Where ϕ_i is the weight of model i, Q_t is the combined model RF power prediction at time t. $e_{i,t}$ and $\hat{P}_{i,t}$ are the error and predicted RF power of model i at time t respectively, m is the number of models combined, n is the observation time index, P_t is the actual RF power at time t while E_t is the combined error at time t.

The equal average method, inverse MSE method and inverse ranking method of obtaining weights of a combine prediction model are given by Equations (4.21), (4.22) and (4.23) respectively.

$$\phi_i = \frac{1}{m} \tag{4.21}$$

$$\phi_i = \frac{\frac{1}{MSE_i}}{\frac{1}{MSE_1} + \frac{1}{MSE_2} \cdots + \frac{1}{MSE_m}} \tag{4.22}$$

$$\phi_i = \frac{\frac{1}{Rank_i}}{\frac{1}{Rank_1} + \frac{1}{Rank_2} \cdots + \frac{1}{Rank_m}} \tag{4.23}$$

Where ϕ_i is the weights of model i. MSE_i and $Rank_i$ are the MSE and rank of model i respectively while m is the total number of models. Rank is an integer in the range

$Rank = 1, 2, \cdots, m$ for the best model, second best model, etc respectively.

4.4.1 Proposed method for obtaining the weights of a combine prediction model

In order to evolve the weights using DE, a novel means of obtaining the range of the weights for each model is proposed. The proposed method combine the advantages of equal weight average method, inverse ranking method and inverse MSE method as shown in Equation (4.24). In equal weight average (EWA) method, all the models are weighted equally irrespective of the individual models performances (e.g. the MSE) and the weight depends only on the number of models used, Equation (4.21), while in inverse MSE (IMSE) method, each model weight is inversely proportional to its MSE, Equation (4.22). Hence in a situation where the MSE of the best model is very small compared to MSEs of other models used, the contribution of the other models in the combined model is very small or insignificant. In some cases, there are some trends that are captured in some of the models which may not be effectively captured in the best model, hence their contribution will enhance the performance or generalization of the combined model. In the inverse ranking method, the means of obtaining the weights did not take into account the relative magnitude of the error of the models i.e. for a given combine model consisting of m models, the weight of the best model and likewise the other models are constant irrespective of the magnitude of their errors. The errors are only used to rank the models, Equation (4.23). The weights of the best model, second best model, etc are all predetermined before the training and validation outcome of the models.

To circumvent the deficiencies of equal average method, inverse MSE and inverse ranking method, a novel bench mark for obtaining the weights of a combine prediction model which combine both the advantages of equal average method and inverse MSE method, and at the same time capturing the effect of the relative magnitude of the validation errors which was not captured by inverse ranking method is proposed as shown in Equation (4.24).

$$\phi_i = \frac{1}{m-1}\left(1 - \frac{MSE_i}{MSE_T}\right) \tag{4.24}$$

Where:

$$\sum_{i=1}^{m} \phi_i = 1, \quad \phi_i \geqslant 0 \tag{4.25}$$

$$MSE_T = \sum_{i=1}^{m} MSE_i \tag{4.26}$$

ϕ_i and MSE_i are the weight and MSE of model i respectively, m is the number of models used while MSE_T is the sum total of all the MSEs of the models. The proposed benchmark shown in Equation (4.24) satisfies the constrain in Equation (4.20), also the higher the MSE of a model the smaller its weight. The use of m in Equation (4.24) smoothed or averaged the weights across the models thereby capturing the contribution

of the individual models without ignoring the effect of the magnitude of their MSE. The bench mark is robust as it is not limited to only MSE but to any empirical loss function (error measurement) that is used to evaluate the performances of the models e.g. mean absolute error, mean absolute percentage error, root mean square error, etc. The bench mark weights may not be optimum but defines the search domain (space) where the optimum or near optimum are likely to be found. The optimised combined model is obtained by fine tuning the ϕ (weights) using differential evolution algorithm variants. The search or decision space of each ϕ is centred within its bench mark, Equation (4.24). In order to maintain the constrain in Equation (4.20), for each generation, the genes of each candidate solution is divided by the sum of the genes (weights).

4.5 Prediction results

The results presented in this section are divided into two subsections i.e. Sections 4.5.1 and 4.5.2. Section 4.5.1 presents the results of RF power prediction using real world RF data captured at different discontinuous time intervals within a period of three years. The RF power prediction models are designed using ANN, SVM, and a combined model consisting of multiple topologies of ANNs and SVMs. Section 4.5.2 consists of the results of spectrum hole prediction using simulated data based on some assumed real world scenarios. As real world spectrum hole data could not be secured, simulated data are used as an alternative. The spectrum hole prediction models are designed using probabilistic ANN. The word probabilistic here referred to the fact that the outputs of the ANN gives the probability of the channels been idle. The results of the spectrum hole prediction using ANN are compared with blind linear and blind stochastic search approaches for various traffic intensities. Most of this section is devoted to RF power prediction, while the spectrum hole prediction is briefly presented as a prove of concept by comparing the use of CI methods and algorithmic approach. See Iliya et al. (2015b) for details of the spectrum hole prediction.

4.5.1 RF power prediction results

The performance of the RF power prediction models implemented are presented in this section. The specification of the ANNs are depicted in Table 4.1.

To minimize the MSE of the ANN and the empirical risk function of the SVR when tested with the test data, the algorithms listed in Section 2.6 were run for 5 independent runs. Each run has been continued with 3000 fitness evaluations. A population of 10 individuals was maintained for 300 generations. After a manual tuning of the parameters, the following parameters were used: The DE is run with mutation constant $F = 0.3$ and crossover probability $Cr = 0.3$, these choice is to enhance the DE moves in order to serve as a global searcher having a high probability of escaping local optimums. The jDE is run with $F_l = 0.1$ and $F_u = 0.9$, and the probability for updated F and Cr are set to $\tau_1 = 0.1$ and $\tau_2 = 0.1$ respectively. The PSO acceleration constants are

Table 4.1: ANN Models Specification

	ANN Models		
	Dedicated one channel	Multiple channels, single output	Multiple channels, parallel output
First Hidden Neurons	5	15	15
Second Hidden Neurons	3	10	10
Output Neurons	1	1	20
Number of Channels	1	20	20

tuned as follows: $C_1 = 0.3$, $C_2 = 0.6$ and $C = 0.3$ for CLPSO. The average mean square error (AMSE) and the standard deviations for 5 independent runs using test data are as shown in Tables 4.2 and 4.3. Table 4.2 presents the results obtained using multiple channels prediction models shown in Figures 4.2 and 4.5, trained for prediction of RF power in 20 subchannels, but one at a time. For this model, the linear kernel SVM, Gaussian RBF kernel SVM, feed forward (FF) ANN and cascaded feed forward (CFF) ANN gives a more promising prediction accuracy on test data as compared to other kernels and ANN topologies implemented, hence their results are the only ones presented in Table 4.2. The linear kernel outperformed the Gaussian RBF with an AMSE of 0.0640 and STD 0.0012, the hyperparameters are evolved using cDE/rand/1/exp. The FF and CFF ANN has the same STD of 0.0009 while their AMSE are 0.0565 and 0.0835 respectively, the FF outperform the CFF.

The RF power prediction results using the parallel model shown in Figure 4.3 trained for simultaneous prediction of RF power in 20 resource blocks is shown in Table 4.4. The model is trained using compact differential evolution variants to evolve the initial weights after which the weights are fine tuned towards the global optimum using backpropagation algorithm (BPA). The best evolved model is obtain using cDE/rand/2/exp with AMSE of 0.0807 and STD of 0.0022. Table 4.4 shows the Wilcoxon test using cDE/rand/2/exp as the reference algorithm. A + indicate that the reference algorithms outperformed the competing algorithm, = means both algorithms are statistical the same while − means that the other algorithm outperformed the reference algorithm Wilcoxon (1945).

The results of the prediction models designed for RF power prediction in only one channel or resource block, using Figures 4.1 and 4.4, is depicted in Table 4.3. In this case, the model designed using exponential kernel emerged as the best with AMSE of 0.0226 and STD 0.0009 while the best ANN topology happen to be FF with AMSE of 0.0215 and STD 0.0010. This study shows that both SVM regression and ANN prediction models are potential good prediction models for this problem even though the ANN outperformed the SVM.

For this problem, the Feed forward (FF) ANN model emerges as the best as compared with other models implemented i.e. feedforward with output feedback (FFB),

cascaded feedforward (CFF) and layer recurrent (LR) ANN. This implies that the feedback information may have been captured through the inputs assigned to the 4 most recent past RF power samples. Comparing the results depicted in Tables 4.2 and 4.3 with similar work detailed in Iliya et al. (2014b); the use of some of the most recent RF power samples as part of the input vector of the ANN, produces a more accurate ANN model with reduced number of neurons.

Since the training algorithms used for evolving the initial weights of the ANN and metaparameters of the SVM are metaheuristics with no means of knowing the best algorithm prior to solving the problem, many variants of the metaheuristics are used. The DE variants outperformed the PSO variants in converging faster to a good solution for evolving the hyperparameters of the SVM and weights of the ANN as depicted in Tables 4.5, 4.2 and 4.3. From the same tables, most of the best solutions were evolved using binomial crossover as compared with the exponential crossover. The PSO will have converge to a good solution if given a longer budget (number of iterations). The performance of the SVM model designed using polynomial kernel is very poor, hence is not included in the results presented. Figures 4.11 and 4.13 shows the predictions using ANN and SVR respectively.

The combined model used in this study consists of six prediction models i.e. three ANNs topologies (FFB, CFF and LR) and three SVM regression models using linear, RBF and exponential kernel. The accuracy of the combine model increases with an average of 46% as compare with the individual models used. Using Equation (4.24) as a bench mark for decision space of the candidate solutions (weights), the DE converges to a good solution within the first 10 generations with jDE/best/2/bin emerging as the best algorithms for this problem. 16 variants of DE algorithms were used to evolve the weights of the combine model, the variation of the combine MSE against number of generation for the best five algorithms are depicted in Figure 4.10. The number of models used were determined experimentally by adding the models one at a time and observing their contribution to the overall accuracy and generalization of the model prediction. It is interesting to note that when the worst model among the six models used, was removed and the other five models are used to design the combine model, the accuracy of the combined model drop as compared to when the six models were used. This reveal that there are some trends or features that characterises the training data that are captured by the worst model. Even though the worst model has the least reference weight as given by Equation (4.24), its contribution impact on the overall combine model performance. The output of the combined model is given by Equation (4.27).

$$Q_t = \sum_{i=1}^{m} \phi_i \hat{P}_{i,t} \tag{4.27}$$

Where Q_t is the combine predicted RF power at time t, ϕ_i is the weight of model i, $\hat{P}_{i,t}$ is the predicted RF power by model i at time t, $i = 1, \cdots, m$ where m is the number of models used. In this case $m = 6$ i.e. 3 ANN topologies and 3 SVM regression models are combined to form the combine model.

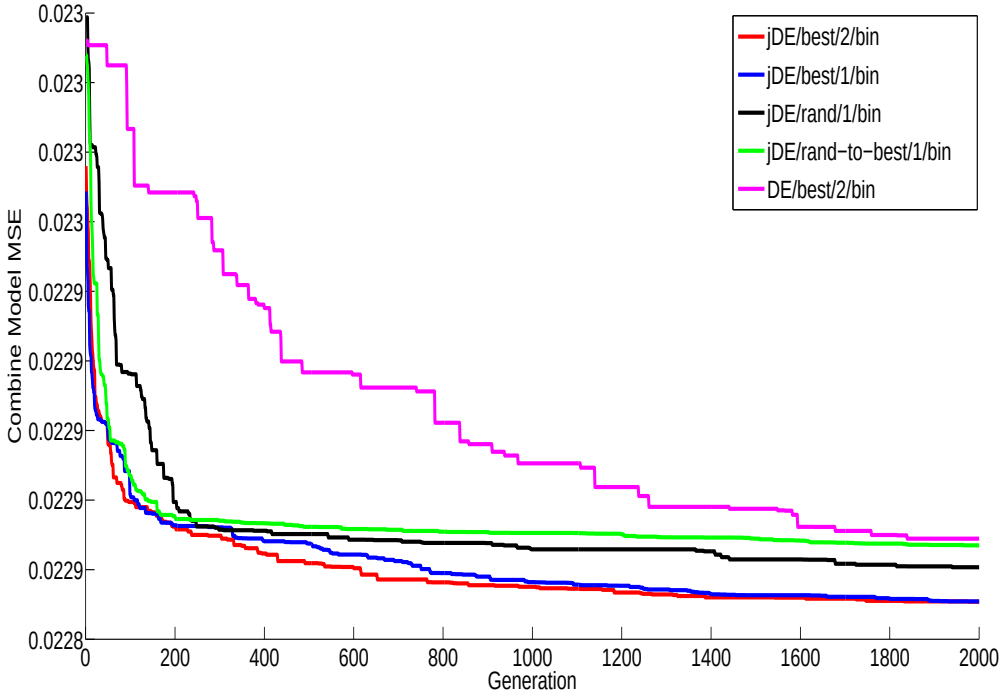

Figure 4.10: MSE of combined model

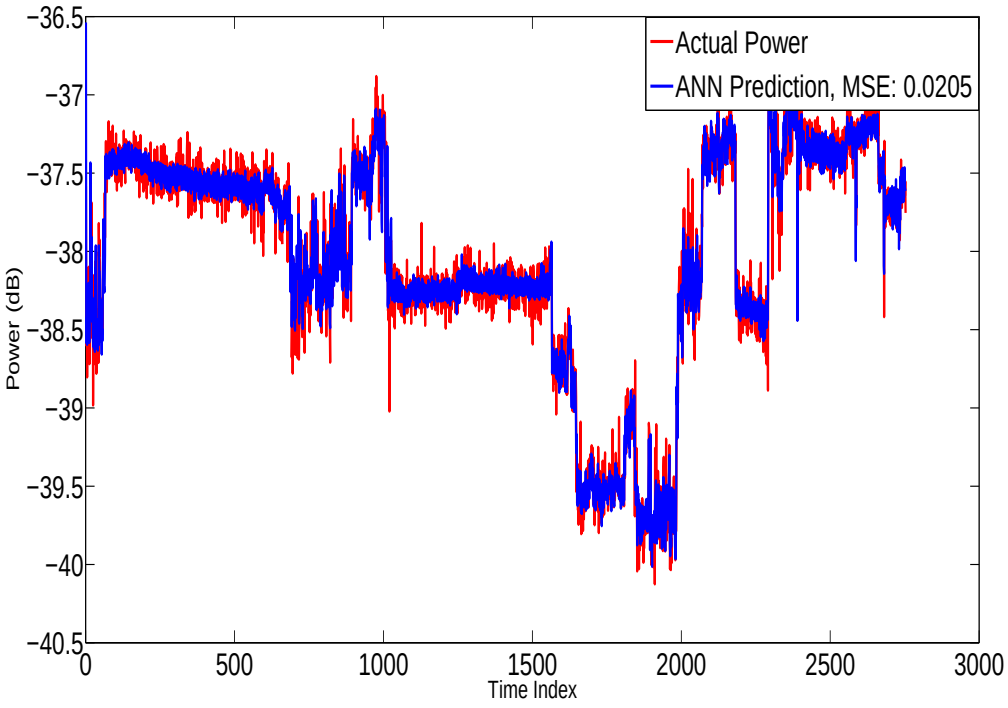

Figure 4.11: One channel ANN Prediction model

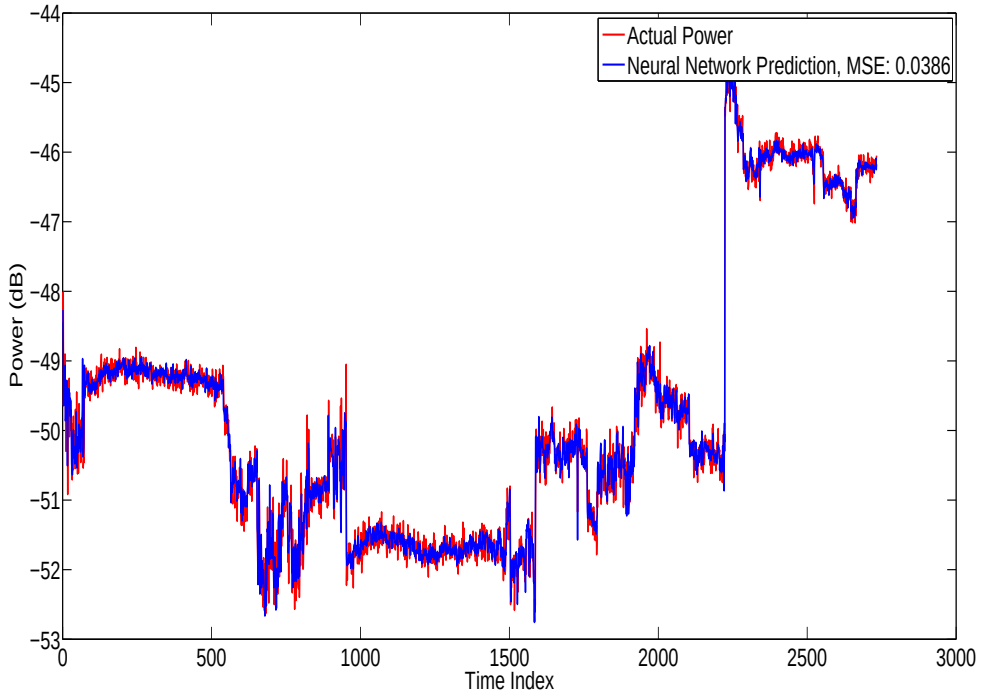

Figure 4.12: Parallel output ANN prediction model

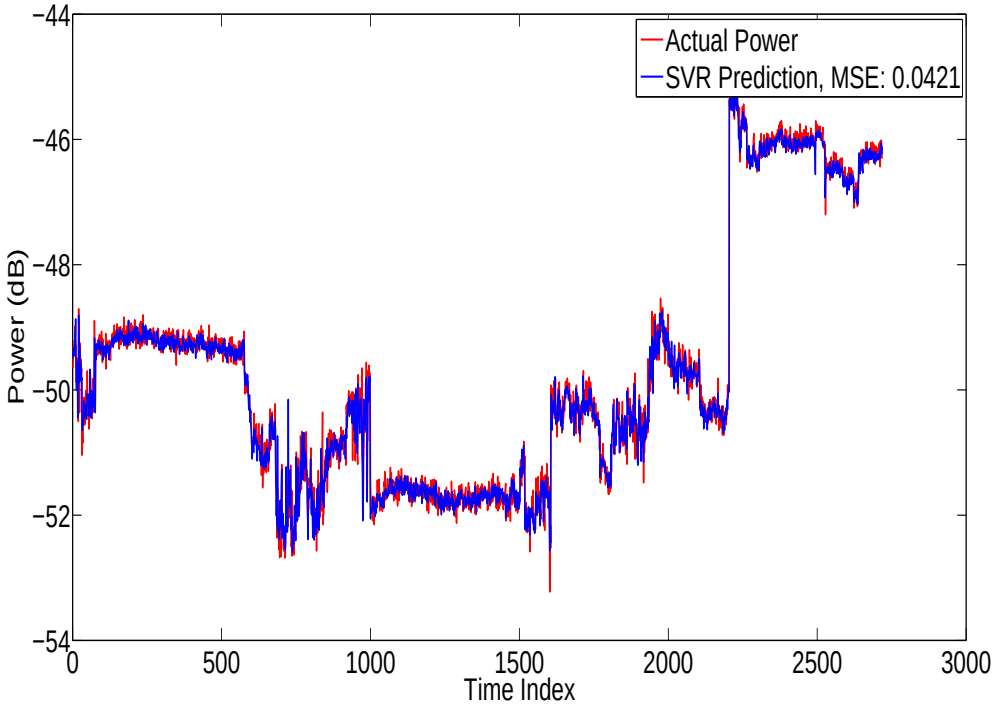

Figure 4.13: Multiple channels SVR Prediction model

4.5.2 Spectrum hole prediction

The results of spectrum hole prediction using probabilistic ANN, blind linear and blind stochastic search are presented in this section. The word probabilistic here referred to

Table 4.2: SVR and ANN results using multiple channel model for 20 channels

Algorithms	Linear		Gaussian RBF		ANN FF		ANN CFF	
	AMSE	STD	AMSE	STD	AMSE	STD	AMSE	STD
PSO	0.1069	0.0018	0.2022	0.0067	0.0806	0.0009	0.0864	0.0030
CLPSO	0.1097	0.0036	0.3703	0.1661	0.0818	0.0025	0.0870	0.0009
S/DE/bin	0.1079	0.0009	**0.1198**	**0.0182**	0.0811	0.0019	0.0887	0.0014
S/DE/exp	0.1081	0.0017	0.1341	0.0198	0.0812	0.0012	0.0848	0.0002
S/DE/bin/S	0.0999	0.0016	0.2482	0.0995	0.0846	0.0024	0.0871	0.0030
S/DE/exp/S	0.1006	0.0012	0.2423	0.1017	0.0819	0.0008	0.0868	0.0014
jDE/rand/1/bin	0.1067	0.0014	0.2048	0.0650	0.0822	0.0008	0.0845	0.0020
jDE/rand/1/exp	0.1090	0.0018	0.2358	0.0471	0.0813	0.0005	0.0865	0.0040
jDE/best/1/bin	0.1055	0.0014	0.1847	0.0169	0.0816	0.0006	0.0873	0.0028
jDE/best/1/exp	0.1057	0.0034	0.2368	0.0298	0.0815	0.0005	0.0881	0.0031
jDE/best/2/bin	0.1065	0.0017	0.1785	0.0209	0.0825	0.0016	0.0863	0.0016
jDE/best/2/exp	0.1090	0.0013	0.2085	0.0134	0.0814	0.0008	0.0848	0.0020
jDE/rand/best/1/bin	0.1069	0.0011	0.1872	0.0182	0.0820	0.0003	0.0865	0.0015
jDE/rand/best/1/exp	0.1071	0.0022	0.2665	0.0862	0.0817	0.0006	0.0849	0.0037
DE/rand/1/bin	0.1062	0.0017	0.1268	0.0251	0.0802	0.0009	0.0864	0.0017
DE/rand/1/exp	0.1063	0.0028	0.2169	0.0305	0.0812	0.0014	0.0848	0.0009
DE/best/2/bin	0.1052	0.0004	0.1734	0.0065	0.0827	0.0019	0.0879	0.0034
DE/best/2/exp	0.1097	0.0004	0.2284	0.0575	0.0818	0.0009	0.0874	0.0019
DE/best/1/bin	0.1048	0.0006	0.1841	0.0293	0.0814	0.0012	0.0848	0.0003
DE/best/1/exp	0.1083	0.0017	0.2219	0.0357	0.0830	0.0009	0.0864	0.0018
DE/rand-to-best/1/bin	0.1068	0.0014	0.2205	0.0559	0.0812	0.0004	0.0852	0.0009
DE/rand-to-best/1/exp	0.1082	0.0006	0.2088	0.0434	0.0826	0.0010	0.0848	0.0002
cDE/rand/1/bin	0.0658	0.0018	0.1274	0.0035	0.0589	0.0019	0.0871	0.0018
cDE/rand/1/exp	0.0660	0.0011	0.1238	0.0025	0.0570	0.0007	0.0864	0.0035
cDE/best/2/bin	**0.0640**	**0.0031**	0.3565	0.3051	0.0575	0.0017	0.0855	0.0039
cDE/best/2/exp	0.0684	0.0061	0.1609	0.0155	0.0576	0.0010	0.0856	0.0029
cDE/best/1/bin	**0.0640**	**0.0012**	0.6258	0.4953	0.0575	0.0009	0.0856	0.0029
cDE/best/1/exp	0.0698	0.0079	0.3395	0.2010	**0.0565**	**0.0009**	0.0861	0.0031
cDE/rand/best/1/bin	0.0680	0.0044	0.2854	0.0425	0.0569	0.0009	**0.0835**	**0.0009**
cDE/rand/best/1/exp	0.0672	0.0036	0.4062	0.0433	0.0575	0.0003	0.0849	0.0006
cDE/rand/2/bin	0.0660	0.0021	0.1281	0.0095	0.0571	0.0009	0.0872	0.0028
cDE/rand/2/exp	0.0671	0.0041	0.1255	0.0036	0.0587	0.0023	0.0891	0.0012

Where: DE referred to differential evolution, bin and exp stand for binomial and exponential crossover respectively (Section 2.6.1), S for Short Distance Exploration algorithm commonly called the S algorithm which is used here as local searcher for super fit DE (Algorithm 2.3), cDE for compact DE (Section 2.6.4), jDE for parametrised DE (Section 2.6.2), PSO for particle swarm optimization while CLPSO signify complementary learning PSO (Section 2.6.5)

the fact that the outputs of the ANN gives the probability of the channels been idle. The ANN was trained to adapt to the radio spectrum traffic of 20 channels and the trained network was used for prediction of future spectrum holes. The input of the ANN consists of a time domain vector of length six i.e. minute (0-59), hour (0-23), date (1-31), day (1-7), week (1-5) and month (1-12). The output is a vector of length 20, each representing the probability of one of the channels being idle. The channels are ranked in order of decreasing probability of being idle. The channels are searched for spectrum hole starting from the channel with the highest probability of being idle, followed by the second and so on. The search stop when the first spectrum hole (idle channel) is found if there exist any. The assumption here is that all the channels have the same noise and quality of service; and only one vacant channel is needed for communication. For the blind linear search, the channels are always search for spectrum hole starting from the first channel, followed by the second and so on. The search stop when a spectrum hole is found or when the last channel is searched. The blind stochastic search is similar to the blind linear search, the only different is that the channels are searched randomly

Table 4.3: SVR and ANN results using one channel dedicated model

Algorithms	RBF		Gaussian RBF		Exponential		Linear		ANN FF	
	AMSE	STD	AMSE	STD	AMSE	STD	AMSE	STD	AMSE	STD
PSO	0.0379	0.0030	0.0725	0.0044	0.0239	0.0004	0.0289	0.0013	0.0240	0.0008
CLPSO	0.0536	0.0179	0.1498	0.0840	0.0924	0.1172	0.0312	0.0004	0.0236	0.0003
S/DE/bin	0.0361	0.0022	0.0644	0.0094	0.0236	0.0002	0.0296	0.0011	0.0237	0.0005
S/DE/exp	0.0373	0.0015	0.0989	0.0288	0.0238	0.0000	0.0300	0.0012	0.0241	0.0011
S/DE/bin/S	**0.0272**	**0.0005**	0.0736	0.0253	0.0291	0.0001	0.0368	0.0007	0.0235	0.0005
S/DE/exp/S	0.0284	0.0012	0.0677	0.0236	0.0288	0.0001	0.0369	0.0000	0.0242	0.0011
jDE/rand/1/bin	0.0378	0.0029	0.0603	0.0109	0.0235	0.0003	0.0287	0.0012	0.0236	0.0005
jDE/rand/1/exp	0.0377	0.0011	0.0956	0.0450	0.0239	0.0001	0.0302	0.0002	0.0235	0.0004
jDE/best/1/bin	0.0366	0.0042	0.0509	0.0009	0.0233	0.0003	0.0289	0.0007	0.0237	0.0005
jDE/best/1/exp	0.0396	0.0000	0.0916	0.0171	0.0242	0.0002	0.0304	0.0002	0.0238	0.0007
jDE/best/2/bin	0.0375	0.0017	0.0583	0.0109	0.0237	0.0001	0.0289	0.0007	0.0239	0.0003
jDE/best/2/exp	0.0389	0.0029	0.0654	0.0093	0.0238	0.0005	0.0302	0.0007	0.0240	0.0011
jDE/rand-to-best/1/bin	0.0378	0.0037	0.0560	0.0054	0.0239	0.0007	0.0288	0.0007	0.0237	0.0002
jDE/rand-to-best/1/exp	0.0425	0.0034	0.1295	0.0782	0.0246	0.0008	0.0301	0.0005	0.0235	0.0004
DE/rand/1/bin	0.0312	0.0000	0.0602	0.0133	0.0238	0.0000	**0.0283**	**0.0013**	0.0235	0.0002
DE/rand/1/exp	0.0386	0.0017	0.0718	0.0009	0.0255	0.0015	0.0302	0.0012	0.0239	0.0008
DE/best/2/bin	00.0369	0.0015	0.0589	0.0091	0.0237	0.0000	0.0295	0.0008	0.0236	0.0003
DE/best/2/exp	0.0416	0.0077	0.0999	0.0070	0.0275	0.0060	0.0300	0.0005	0.0239	0.0004
DE/best/1/bin	0.0356	0.0006	**0.0506**	**0.0003**	0.0239	0.0006	0.0289	0.0009	0.0238	0.0004
DE/best/1/exp	0.0395	0.0016	0.1617	0.0749	0.0242	0.0001	0.0300	0.0012	0.0236	0.0003
DE/rand-to-best/1/bin	0.0400	0.0017	0.0569	0.0107	0.0239	0.0002	0.0292	0.0008	0.0237	0.0005
DE/rand-to-best/1/exp	0.0422	0.0022	0.1500	0.0626	0.0246	0.0001	0.0302	0.0003	0.0236	0.0004
cDE/rand/1/bin	0.4055	0.0780	0.0614	0.0198	0.0243	0.0006	0.0301	0.0005	0.0236	0.0002
cDE/rand/1/exp	0.1950	0.0964	0.1333	0.0916	0.0298	0.0111	0.0313	0.0012	**0.0215**	**0.0010**
cDE/best/2/bin	0.0508	0.0174	0.2043	0.1437	0.0350	0.0181	0.0303	0.0003	0.0237	0.0003
cDE/best/2/exp	0.0416	0.0078	0.2114	0.0788	0.0437	0.0190	0.0313	0.0005	0.0247	0.0024
cDE/best/1/bin	0.0474	0.0015	0.2670	0.0385	0.0266	0.0041	0.0310	0.0007	0.0239	0.0005
cDE/best/1/exp	0.0511	0.0035	0.2115	0.1355	0.0323	0.0015	0.0323	0.0015	0.0238	0.0003
cDE/rand-to-best/1/bin	0.0477	0.0045	0.2062	0.0856	0.0243	0.0001	0.0314	0.0007	0.0240	0.0013
cDE/rand-to-best/1/exp	0.0507	0.0037	0.1346	0.0806	0.0239	0.0014	0.0306	0.0008	0.0236	0.0004
cDE/rand/2/bin	1.6718	0.0000	0.1937	0.0946	0.0248	0.0008	0.0306	0.0005	0.0236	0.0003
cDE/rand/2/exp	0.2087	0.0302	0.1912	0.0625	**0.0226**	**0.0009**	0.0313	0.0015	0.0237	0.0005

Table 4.4: CFF ANN results for 20 channels using parallel output model

Algorithms (cDE + BPA)	AMSE	STD
cDE/best/2/bin	0.0854	0.0089+
cDE/best/2/exp	0.0808	0.0020=
cDE/best/1/bin	0.0849	0.0061+
cDE/best/1/exp	0.0811	0.0047+
cDE/rand/to/best/1/bin	0.0812	0.0033+
cDE/rand/to/best/1/exp	0.0823	0.0034+
cDE/rand/2/bin	0.0816	0.0029+
cDE/rand/2/exp	**0.0807**	**0.0022**
cDE/rand/1/bin	0.0809	0.0017=
cDE/rand/1/exp	0.0826	0.0037+
BPA(constant learning rate)	0.1345	0.2029+
BPA (varying learning rate)	0.2508	0.0332+

Table 4.5: Best SVR model parameters for Table 4.3

Kernel	Algorithms	MSE	C	ε	γ or σ
RBF	S/DE/bin/S	0.0267	13.73	0.000142	6.31
Gaussian RBF	cDE/rand/1/bin	0.0386	7.62	0.000333	1.49
Exponential	cDE/rand/2/exp	0.0217	22.28	0.000228	7.57
Linear	jDE/bin	0.0274	15.97	0.000021	-

not serially. The word blind here is used to indicate that there is no any evaluation metric that is use to justify the search direction (order).

The simulation data were obtained as follows: Is a well known fact that majority (i.e. over 50%) of the assigned or licensed spectrum are not efficiently utilized with respect to time, frequency and spatial domain, Communication (2003), OFCOM (2007). The simulation data were generated using different traffic intensities which represent the percentage utilization of the channels. Since the spectrum activities under study are man-oriented, sometimes it follows certain scenarios (trends) and some times stochastic in nature. Some spectrum may be busier during the day time than night or at other time instances. While some may be heavily used during weekends than working days or at certain dates, seasons or during some special events Bacchus et al. (2010), Qaraqe et al. (2009). As a result of economic and other factors, some spectra are more utilized in urban areas than rural areas which also vary in developed and developing countries; while some are only confined to certain geographical locations e.g. the aeronautical spectrum. Considering these factors, weights are used to represent the percentage of idle period for each channel with respect to six elements i.e. minute, hour, date, day, week and month. These weights give information about the channels utilization or the traffic intensity. To account for the contribution of each of these six elements, crisp values were obtained by finding the average of the weights. These crisp values represent the proportion of idle state of the channel. Random 1s and 0s were generated using a normal random number generator according to the traffic intensity. Figures 4.14 to 4.18 shows the results expressed in percentage against number of channels searched before the first spectrum hole (idle channel) is found. The total number of time instances spectrum opportunity is demanded during the simulation is 7440. The legends of Figures 4.14 to 4.18 for the various traffic intensities, referred to the percentage of the total number of search (7440) at which the referenced search method found a spectrum hole (idle channel) at first search. For a traffic intensity of 0.3262 (32.62%), out of the 7440 search, the ANN after ranking the channels starting from the one with the highest probability of being idle, 97.87% (7281) of the total search the ANN found spectrum hole at first search while that of the blind linear and blind stochastic search are 62.88% and 65.44% respectively as shown in Figure 4.14. As the traffic intensity increases, more channels become busy (occupied) and the chances of locating an idle channel at first search becomes difficult, yet even at a high traffic intensity of 62.62%, the ANN exhibit an amazing level of intelligence such that 65.05% of the total search, spectrum holes were found at first search. The linear and stochastic search achieve 33.27% and 35.38% respectively for the same traffic intensity of 62.62% as shown in Figure 4.18. The results for other traffic intensities are presented in Table 4.6. The use of ANN does not only speed up idle channel selection, but it also save signal processing energy that will have be wasted in sensing busy channels. The probabilistic ANN used in this research outperform the ANN used by Baldo et al. (2009) for dynamic channel selection.

Table 4.6: Percentage of spectrum hole (idle channel) found at first search for various traffic intensities

Traffic Intensity %	Probabilistic ANN %	Blind Linear %	Blind Stochastic %
32.62	97.87	62.44	65.44
37.62	93.53	57.25	59.51
42.62	87.05	51.97	55.67
52.62	78.21	42.47	45.92
62.62	65.05	33.27	35.38

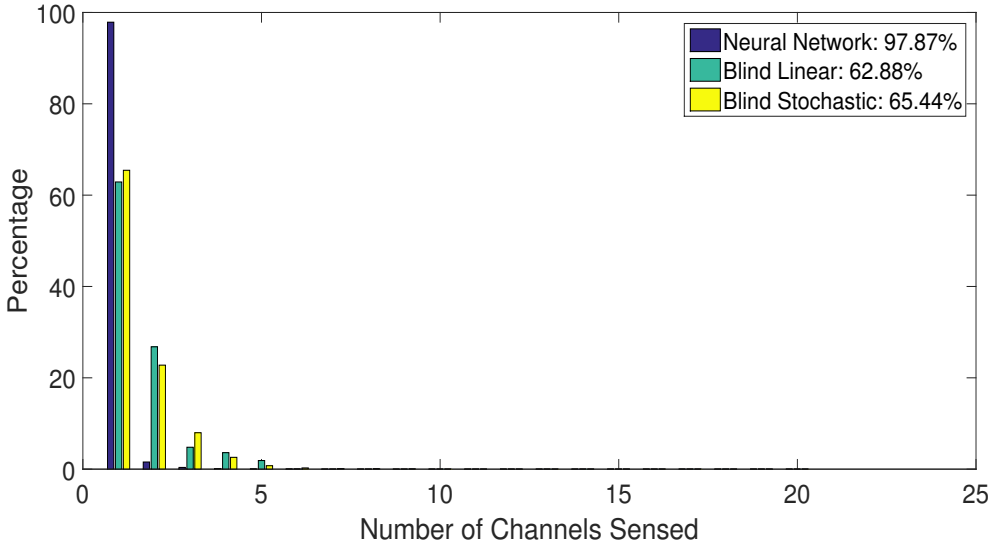

Figure 4.14: Percentage of number of channels searched before the first spectrum hole is found for traffic intensity of 32.62%

4.5.3 Conclusion

The results of the RF power prediction of both the ANN and SVM regression models using test data shows that both models are well generalized and stand out as a good choice for this problem. For the model designed for prediction of RF power in one channel, the ANN outperformed the SVM with average mean square error (AMSE) of 0.0215 while that of SVM is 0.0226, likewise for the multiple channel, single output prediction model, the AMSE of ANN is 0.0565 while that of SVM is 0.0640. Linear kernel emerged as the best for the model designed for prediction of RF power in single output, multiple channels model and exponential kernel for one channel dedicated model. For the ANN multiple channels parallel output model, the AMSE of the best model is 0.080. Thus among the three models designed, the dedicated model designed for prediction of RF power in one channel achieved the best performance with AMSE of 0.0215. But this approach is not practically economical for use in broadband prediction since all the channels or resource blocks will have an assigned dedicated prediction model. Considering the computing demand, and the memory and power (battery power) constraint of cognitive radios, the recommended model is the model designed for prediction of RF

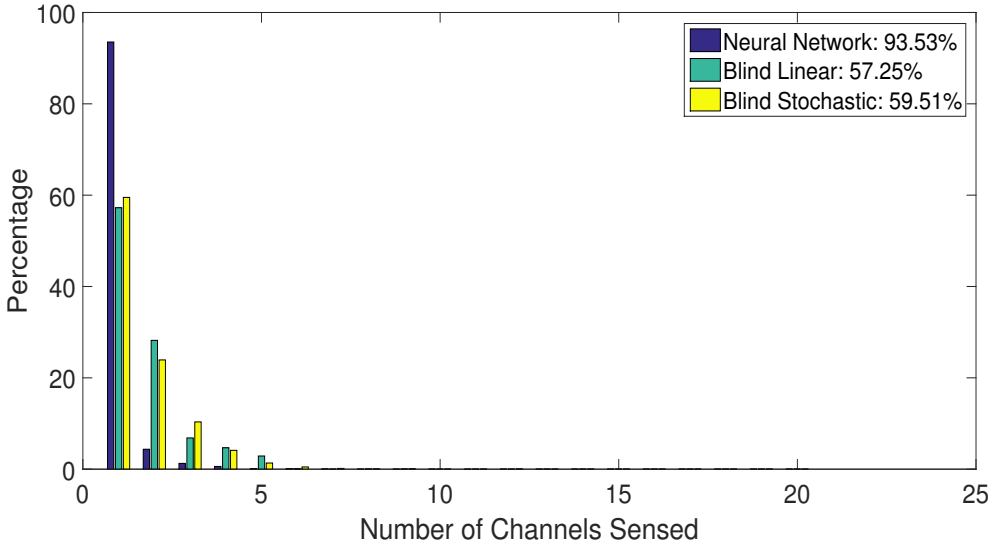

Figure 4.15: Percentage of number of channels searched before the first spectrum hole is found for traffic intensity of 37.62%

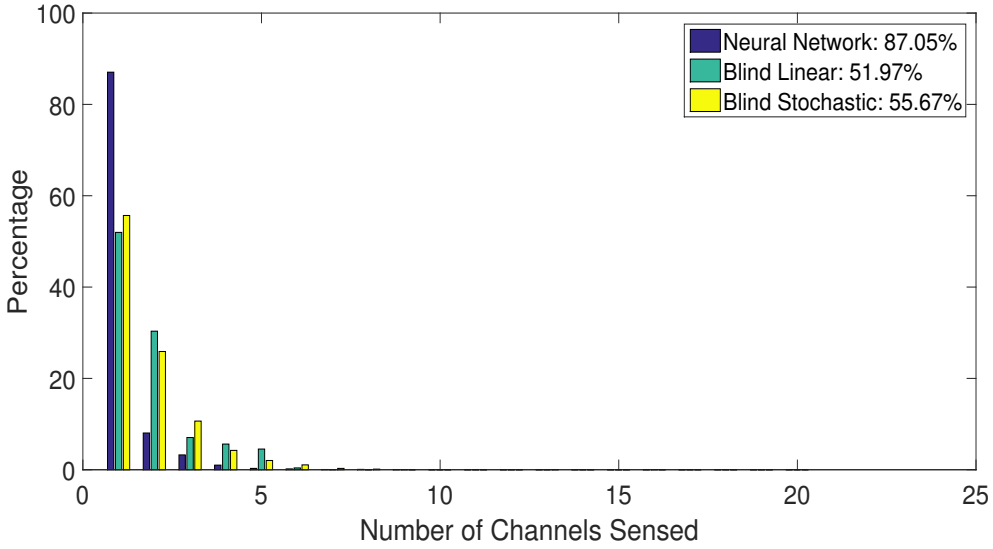

Figure 4.16: Percentage of number of channels searched before the first spectrum hole is found for traffic intensity of 42.62%

power in multiple channels but one at a time. This model has an additional input for channel identification and only one output. The compact DE (cDE) seems to be the most promising for this problem. The a priori knowledge of the RF power resulting from either communication signals, noise and/or interferences, is not only applicable to cognitive radio network, but in any wireless communication system for noisy channels avoidance. The location input of the proposed prediction models were not active since the data used in this study were captured in one geographical location and at the same height. Feature work should involve measurements conducted simultaneously in more than one geographical locations at different heights within the same period of time.

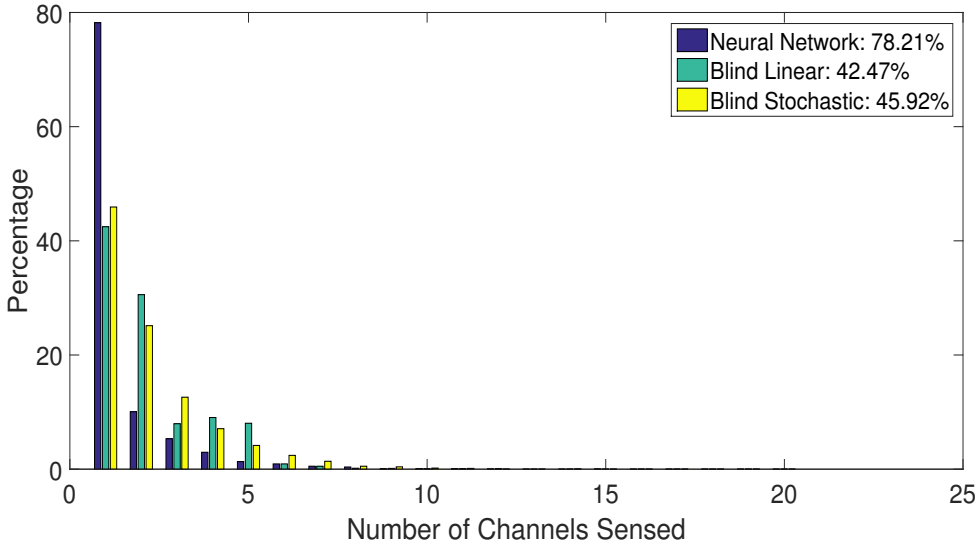

Figure 4.17: Percentage of number of channels searched before the first spectrum hole is found for traffic intensity of 52.62%

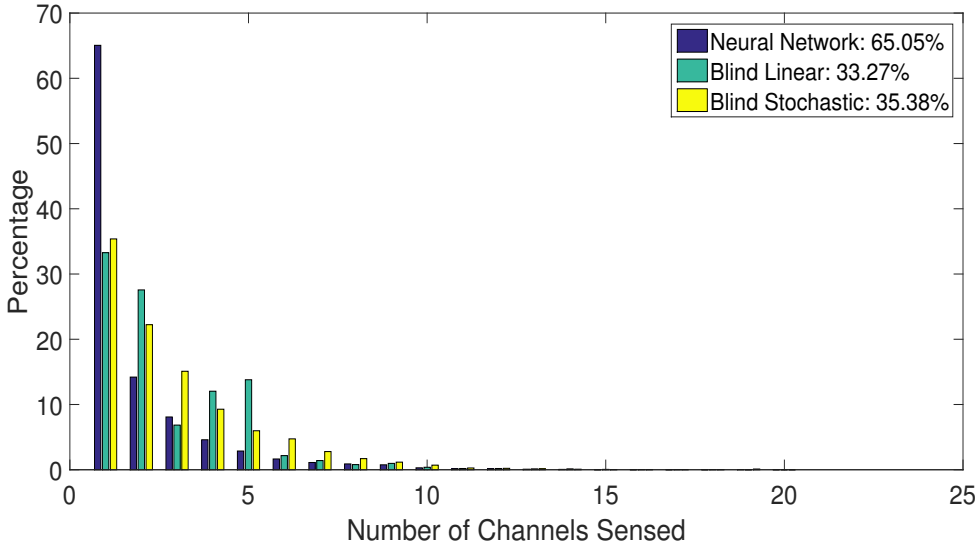

Figure 4.18: Percentage of number of channels searched before the first spectrum hole is found for traffic intensity of 62.62%

This will help in the evaluation of the proposed prediction models, and also provide a wider view of spectrum occupancy estimate that can facilitate cognitive radio network implementation.

The skills and knowledge acquired in signal processing and CI from this research were used to address some problems related to medical field which are presented in the next chapter.

Chapter 5

Application of Computational Intelligence in Speech Therapy

5.1 Introduction

Though the main focus of this research is on the application of computational intelligence (CI) and machine learning in cognitive radio network for effective spectrum utilization; within the period of this study, the knowledge and experiences acquired from the PhD research in signal processing and CI have also been used to address certain problems related to the medical field. This include the application of artificial neural network (ANN) and support vector machine (SVM) for segmentation of impaired speech signal. The segmentation is the first phase of a research that aim at developing and artificial speech therapist which can serve as a self training device capable of providing audiovisual feedback progress report to speech impaired patient. From this first phase of the research, two IEEE conference papers were published, Iliya et al. (2014d;c) and one journal published by International Journal of Neural Systems (IJNS), Iliya and Neri (2016).

The remaining sections are centred on impaired speech segmentation using ANN and SVM introduced in Section 5.2. Background information and related works in speech processing are presented in Section 5.3. This is followed by Section 5.4 where the methods adopted for extracting the speech features needed for the identification of the various sections of the impaired speech signal are discussed. A robust user friendly GUI tool was developed using matlab for normal and impaired speech signal processing that contain both algorithmic signal processing methods and computation intelligence framework approaches as presented in Section 5.4.7. Computational intelligence multi-class classification models were implemented using ANN and SVM as presented in Sections 5.5 and 5.6 respectively. Section 5.6.1 present a novel approach for evolving the metaparameters and weights of the SVM classifier using the proposed hybrid ad-hoc nested algorithms consisting of metaheuristics and convex optimization algorithm. This is the extension of the algorithms used for training of the SVM regression models for

RF power prediction discussed in Chapter 4, now used to train a SVM classifier for impaired speech segmentation. The results and summary of the findings are presented in Sections 5.7 and 5.8 respectively.

5.2 Impaired Speech Segmentation Using ANN and SVM

This section presents a computational intelligence based signal processing technique for segmenting short impaired speech utterances into silent/noise, unvoiced and voiced sections. The proposed technique identifies and captured those points of the impaired voiced speech where the spectrum becomes steady. In contrast to standard speech recognition techniques, this technique aims at detecting that limited section of the speech which contains the information about the potential impairment. This segment of the speech is of interest to the speech therapist as is related to the dynamics of the speech organs (tongue, lips, jaw, glottis, vocal track, etc) responsible for the wrong articulation. This segmentation process is part of a system for deriving musculoskeletal articulation data from disordered speech utterances, in order to provide training feedback for people with speech articulation problem. To address this problem, two segmentation models were developed using support vector machine (SVM) and artificial neural network (ANN) for identification and capturing of the various sections of the disordered (impaired) speech signals. Four neural network topologies were combined to form a neural network mixture model. The approach implements a novel and innovative means of evolving the hyperparameters of the SVM model. The SVM is trained using a hybrid ad-hoc nested algorithm whose outer layer consist of different variants of population based and compact metaheuristic optimization algorithms given in Section 2.6, while the inner layer is a convex optimization algorithm. Several metaheuristics are used and their results compared leading to the conclusion that some variants of the compact differential evolution algorithm appears to be well-suited to address this problem. Numerical results show that the support vector machine modelled using radial basis function kernel is capable of effective detection and isolation of the sections of speech that are of interest to a therapist. The best performance has been achieved when the system is trained by the nested algorithm. A population based approach displays the best performance for the isolation of silence/noise sections, and the detection of unvoiced sections. On the other hand, a compact approach appears to clearly be well-suited to detect the beginning of the steady state of the voiced signal.

The performances of the two proposed segmentation models were compared with one of the famous modern segmentation model based on Gaussian Mixture Model Brognaux and T. (2016) and are found to be superior. The results presented in this chapter are the outcomes of the first phase of the research which aim at developing an artificial speech therapist capable of providing reliable text and audiovisual feed back progress report to the patient.

5.3 Background information and related work in speech processing

Articulation refers to the movement of the tongue, lips, jaw, glottis, vocal track and other speech organs (articulators) to produce speech sounds. The shape of the vocal tract can be controlled in a number of ways by modifying the position or movement of the speech articulators. Phonetic disorders are often attributed to improper articulation. In some cases, speech disorder is often associated to inability of more than one articulators to articulate properly, thus in order to identify the causes of the disorder, and to provide reliable solutions and feedback progress report to the patient, there is need for proper segmentation of the impaired speech signals, Penaloza et al. (2013; 2015). The segmentation of impaired speeches from different patients of different races, and gender can be complex and demanding; computational intelligence and machine learning techniques can be applied to this scenario. This is the first stage in a system, now in development, that aims at predicting musculoskeletal articulation of the vocal tract and other speech articulators from disordered speech. Instead of a patient arranging for frequent physical meetings with a speech therapist for training, articulation prediction models could be used within systems that help speech disorder patients train themselves, in addition to the help they receive from professional human therapists. The present study follows on from an earlier study where a library of 3-dimensional representations of vocal articulation were demonstrated as a useful training resource, Cornelius et al. (2011). With articulation prediction, it is possible to animate these representations precisely according to detailed nature of the disorder, and to identify the articulators that are responsible for each segment (section) of the impaired speech.

For a training process to be effective, the prediction models should respond to small changes in the disordered utterance so that the patients can know whether they are progressing in a useful way. Therefore the speech analysis component should be sensitive to small changes, and also able to characterize the articulations accurately across a wide variety of voice types. By contrast, speech recognition systems are required only to find the most likely text matching a sound, not to evaluate the quality of utterance. Speech recognition is aided by the context present in a string of sounds. In the proposed disordered speech training system the patient is requested to speak only short utterances. The current study uses sound of the form vowel - consonant - vowel e.g. ASA. The leading vowel sound is added to make it easier for the patient to pronounce the following part. In order to extract structured information from the sound, the utterance is first segmented by isolating the silent and/or noise sections. After isolating the voiced activity sections of the impaired speech, the unvoiced and voiced sections of the final consonant - vowel are captured, and also the point where the vowel spectrum becomes steady was identified. The first point where the spectrum becomes steady is very significant as it likely to mark the end of the most informative part of the speech that contain details of the articulation problem which may be useful to speech therapist. Within the steady state, there is no much or new spectral information

about the articulation problem. To evaluate an automated segmentation it is compared with a segmentation prepared by hand that is considered optimal. Artificial neural network and SVM are used to optimize over the parameter spaces of the segmentation methods. A variety of derived signals are used in the segmentation process. Performing a good segmentation is not straightforward, hence the application of ANN and SVM.

While the current work is focused on articulation analysis, related work in speech recognition is briefly reviewed here. There are different types of Computational Intelligence (CI) and machine learning techniques that can be used in speech signal processing such as genetic algorithms, fuzzy logic system (FLS), neural network, hidden Markov model (HMM), Gaussian mixture model (GMM), differential evolution, swamp intelligence, etc. Some for learning and speech recognition, some for optimization of certain processing parameters while others for decision making. Speaker recognition using ANN and hidden Markov model (HMM) was proposed, Dey et al. (2012), in which the power spectral density of speech signals obtained from five people were used in training the ANN and the HMM for identification of the five participants. The models mapped a given speech utterance to non or one of the five participants for security purposes. In order to analyse speech articulation a more detailed approach is proposed than those used for speech recognition. Each section of the sound should first be identified followed by the analysis of the dynamics within each tracked. The work presented in this thesis is concerned with the first identification part.

Phones and words recognition using deep neural network (DNN) was proposed, Siniscalchi et al. (2013). The proposed scheme was found to increase the accuracy of phones and words recognition significantly over the conventional Gaussian mixture model (GMM), by arranging the DNNs in a hierarchical structure to model long term energy trajectory. Environmental sound recognition using matching pursuit (MP) algorithms was proposed by Chu et al. (2009). The proposed method was used to recognize environmental sounds such as rain and chirpings of insects, vehicles, etc which are noise-like in nature, typically with a broad flat spectrum and some times may include strong temporal domain signatures. The MP was used to obtain effective time-frequency features to supplement the mel-frequency ceptral coefficients (MelFCC) thus yielding higher recognition accuracy. Probabilistic artificial neural (PANN) was used for recognition and classification of vowels signals (a, e, i, o, u) into their respective classes, Lim et al. (2000). Vowels sounds from 200 participants comprises of both male and female subjects from different races were used in training the PANN; an average accuracy of 98.4% was claimed to be recorded. From the aforementioned, it is obvious that almost all the speech recognition systems aim at mapping a giving speech signal to either a speaker Prithvi et al. (2014), an alphabet, number, word, sentence or text with no attention given to the critical issues that characterized the speech from people with speech or articulation impairment. This research is the first stage toward the design of an artificial speech therapist focusing on the identification of some of the various sections of disordered speech signals.

5.4 Speech Features

The aim of the segmentation is identification of the limited sections of the speech which is likely to contain the information about the speech impairment. In order to achieve this aim, the speech signal has been analysed and five speech features consisting of both time and spectra (frequency) features were extracted. These features are: energy, zero crossing count rate (ZCCR), linear predictive coding (LPC) coefficients, autocorrelation coefficients at unit sample delay Dash et al. (2012), and mel frequency cepstral coefficients. Since the steady states of interest are those within certain voiced sections (second vowel sound) of the disordered speech, some of the features that characterize normal voiced speech were exploited for identification of the steady states.

5.4.1 Direct Current (DC) removal and pre-emphasis

The preprocessing of the signal consists of the removal of the DC offset. This DC offset was removed by subtracting the mean of the speech signal from each of the speech samples as shown in Equation (5.1). The DC offset removal is vital for correct determination of ZCCR of each frame, as explained in Section 5.4.3. The frame size (window) in this research is 10ms for a sampling frequency of 16 KHz. The choice of 10ms is based on the fact that speech sound are slow time varying signals which are often stationary within a time window of $[5-25]$ ms. The stationarity here implies that the formants, and the position and shape of the articulators remain unchanged within a time frame of $[5-25]$ ms, Wolfel and McDonough (2009). The DC offset removal is done over the entire speech signal and is expressed by the following formula:

$$s(n) = x(n) - \frac{1}{N}\sum_{n=0}^{N-1} x(n), \quad n = 0, 1, 2....N-1 \tag{5.1}$$

where x is the speech with DC offset, s is the speech without DC offset, n is the time index and N is the number of samples of the speech signal.

After the DC removal, the high frequency components of the speech were emphasized using a high pass finite impulse response (FIR) filter shown in Equation (5.2).

$$H(z) = 1 - 0.95z^{-1} \tag{5.2}$$

where z is the complex variable of the filter, see Lawrence and Gold (1975) for details.

The high frequency components were emphasized (boosted) in order to compensate for the de-emphasis (attenuation) of the high frequency components that took place in the glottis during voiced speech production, Li and Douglas (2003). The pre-emphasis flattens the speech signal spectrally thus reducing the large dynamic range of the speech spectrum by adding a zero to the spectrum. The flattening enhances the modelling of the formants using LPC.

5.4.2 Short time energy

The short time energy is the energy computed for each frame using Equation (5.3). This energy is defined as the sum of the square of the speech samples within a frame, Chu et al. (2009):

$$E_i = \sum_{n=0}^{L-1} s_i(n)^2 \tag{5.3}$$

Moreover, the short time power is the energy per short time unit, Equation (5.4):

$$P_i = \frac{E_i}{L} \tag{5.4}$$

where P_i and E_i are the short time power and energy of frame i respectively, s_i is speech samples of frame i and L is the number of samples per frame.

Regarding the expected energy values, the energy of silent sections is obviously less than the energy of any form of speech. It might occur that the background noise can be very high thus resulting into a very high energy. Voiced speech sections are generally characterized by higher energy values than unvoiced speeches, Saha et al. (2005).

5.4.3 Zero crossing count rate

The zero crossing count rate (ZCCR) is the number of times the speech signal changes sign. This feature is the measure of the frequency at which the energy is concentrated in the signal spectrum.

$$ZCCR_i = \sum_{n=0}^{L-2} \frac{|sign(s_i(n)) - sign(s_i(n+1))|}{2} \tag{5.5}$$

where $ZCCR_i$ = ZCCR of frame i and $s_i\,(n)$ are the pre-processed speech samples from Equation (5.1). It may be noted that two consecutive samples with opposite sign (i.e. a zero crossing) result into adding 2 to the numerator, hence the division by 2.

The ZCCR is computed frame by frame hence it is referred to as the short time ZCCR. Voiced speech is produced due to excitation or vibration of the vocal tract by periodic flow of air at the glottis and often have low ZCCR compare to unvoiced speech Radmard et al. (2011). Silence has lower ZCCR than unvoiced but quite comparable to voiced speech. The ZCCR is calculated using Equation (5.5) after removing the DC offset Aida–Zade et al. (2008).

5.4.4 Normalized autocorrelation coefficient at unit sample delay

This index is the measure of the correlation between two consecutive or adjacent speech samples and is defined as:

$$NC_i = \frac{\sum_{n=1}^{L} s_i(n)s_i(n-1)}{\sqrt{\sum_{n=1}^{L} s_i(n)^2 \sum_{n=0}^{L-1} s_i(n)^2}} \tag{5.6}$$

where NC_i is the normalized autocorrelation coefficient of frame i at unit sample delay and $s_i\,(n)$ are the pre-processed speech samples from Equation (5.1).

Voiced speech is characterized by a high concentration of low frequency energy. Hence, adjacent samples of voice speech signal are highly correlated. Consequently, for voiced speech this parameter is close to 1. In unvoiced speech, the concentration of high frequency energy makes adjacent samples to be uncorrelated, hence the normalized autocorrelation coefficient at unit sample delay (NACUSD) is close to zero or negative, QI et al. (2004).

5.4.5 Steady states detection

Steady states refer to the sections of speech signal with no variation in the frequency domain. In order to estimate the beginning of the steady state, in this study a time windows of 0.08s (8 frames) are observed. In other words, if there is no variation in the formants within 0.08s, the signal is considered to be a steady state. This correspond to the disordered speech section just after the unvoiced section in which the average rate of change of the formants is fairly constant within a given threshold for at least 0.08s (8 frames).

In order to locate the steady states, each speech frame is windowed using Hamming window to reduce spectral leakage due to discontinuities, and to enhance the convergence of the Durbin's algorithm for computation of the LPC coefficients, Perez-Meana (2007). LPC was used to obtain 13 LPC coefficients for each windowed frame. LPC modelled the vocal track as an all poles filter, Equation (5.7) from which the formants frequencies can be inferred.

$$H(z) = \frac{S(z)}{U(z)} = \frac{b_0}{1 - \sum_{i=1}^{p} a_i z^{-i}} \tag{5.7}$$

where b_0 is the excitation gain and the coefficients $a_1, a_2, \cdots, a_p$ are the weights of past p samples and z is the complex variable of the transfer function.

The main idea in LPC is that a speech sample $s(n)$ at time n can be modelled as a linear combination of its past p samples, Equation (5.8):

$$s(n) = b_0 u(n) + \sum_{i=1}^{p} a_i s(n-i) \tag{5.8}$$

where $u(n)$ is the excitation signal.

If the z-transform is applied to Equation (5.8), the following expression is obtained:

$$S(z) = b_0 U(z) + \sum_{i=1}^{p} a_i S(z) z^{-i} \tag{5.9}$$

The first, second and third formants frequencies were obtained from the resonance frequency of the vocal tract filter model in Equation (5.7). Since spectral transition plays important role in this research, the first derivative of the formants frequencies was obtained and the sum of the absolute values of the formants derivatives (SAVFD) is computed.

The derivatives obtained directly from the formants frequencies are naturally quite noisy. In order to filter the noise, these derivatives were obtained via polynomial approximation. A second order polynomial is fit into the trajectory of each of the three formants, the derivative is obtained using Equation (5.10):

$$\delta(n, m) = \frac{\sum_{k=-q}^{q} f(n, m+k)k}{\sum_{k=-q}^{q} k^2} \tag{5.10}$$

where $f\,(n, m+k)$ is the formant at position $n = 1, 2$ or 3 (i.e. $1^{st}, 2^{nd}$ or 3^{rd} formant) of frame $m = 1, 2, 3, 4, \ldots$, and δ is the derivative. The parameter q has been set to 3 in this research as it gives a good approximation and not too long delay. The fitting width $2q + 1$ introduces q frames of delay. The sum of the δ functions is the SAVFD, Padmanabhan (2003), Lawrence and Gold (1975).

5.4.6 Mel Frequency Cepstral Coefficients (MelFCC)

The MelFCC is a vector of amplitude of the coefficients of the inverse discrete Fourier transform (IDFT) obtained from each frame of the speech. This vector modelled the human auditory perception of a speech i.e. each amplitude of the frequency (bin) is modelled or compressed according to the sensitivity of the human ear to that frequency. In this study 13 MelFCC are extracted from each frame. 13 are used as it appeared to offer a high detection accuracy of the phenomenon while further increase do not enhance the accuracy of the model.

In order to obtain the MelFCC vector, the speech is pre-emphasised and the spectral amplitude of the discrete Fourier transform (DFT) bins of the speech frames are obtained. The amplitude are passed through a triangular filterbank (integrator) followed by a logarithmic amplitude compression. Inverse discrete Fourier transform is then applied to it and the resulting amplitude of the spectrum are the MelFCCs. The triangular filter banks are spaced using the mel scale which modelled human auditory perception as shown in Equation (5.11).

$$Mel\,(f) = 2595 \log\left(1 + \frac{f}{700}\right) \tag{5.11}$$

where f is the frequency of the signal, log is the decimal logarithm. This scale is approximately linear upto 1KHz after which the sensitivity of human ear to sound (speech) is model as shown in equation (5.11), Sahidullah and Saha (2012).

The speech features described above i.e. the Power (P_i), NACUSD, ZCCR, SAVFD, MelFCC are used as inputs for the two neural models described in Sections 5.5 and 5.6.

5.4.7 Analysis tool

The signal processing, analysis and design of the CI models were done using the matlab environment. A user friendly graphical user interface (GUI) was developed to ease model design. The recordings used for the training and testing of the models were done in English language, to increase the robustness of the tool and to cover a wide range of languages and voice types, the tool developed allow the user to train the model using any language of their choice. The tool can also be used for visualization of many speech features and to observe their dynamics as related to the various sections of the impaired speech. This is very important as it enable the choice of the most relevant features for training and prediction. Some of the features that can be visualized includes: Root mean square (RMS) value, wave, energy, power, spectrogram, LPCs, MelFCCs, formants, power spectral density (PSD), periodogram, single step autocorrelation, zero crossing count rate, etc. The tool include facilities to play the audio recordings. Three different segmentation CI models consisting of different topologies are included in the tool. These are: ANN, SVM and Gaussian mixture model (GMM). The main GUI is as shown in Figure 5.1, the SVM and Neural Network button open the GUI for SVM and ANN as depicted in Figures 5.2 and 5.3 respectively. Figures 5.2 and 5.3 are only used for training, after the training, the best models will be saved and used for prediction using the main GUI, Figure 5.1.

Some of the speech features from a male subject are as shown in Figures 5.4 to 5.7. This speech consist of vowel-consonant-vowel sound i.e. 'asa' repeated five times as can be seen from the wave subplots. Figure 5.4 and 5.5 display the spectrogram, waveform, the first four mel-frequency ceptral coefficients (MelFCCs), and the power spectral density (PSD) of the impaired speech with silent/noise sections and the one with silent/noise removed respectively. It can be seen that the MelFCCs follow certain repetitive pattern within the various sections of the speech signal. In the same vein Figure 5.6 and 5.7 shows the waveform, RSM value, normalized zero crossing count rate (ZCCR) and normalized autocorrelation coefficient at unit sample delay (NACUSD)of the same speech. It is evident that the ZCCR and NACUSD are approximately close to 1 and -1 respectively within the unvoiced segments of the speech. Hence this tool can be used for selection of the most appropriate speech features for a given impaired speech segmentation model.

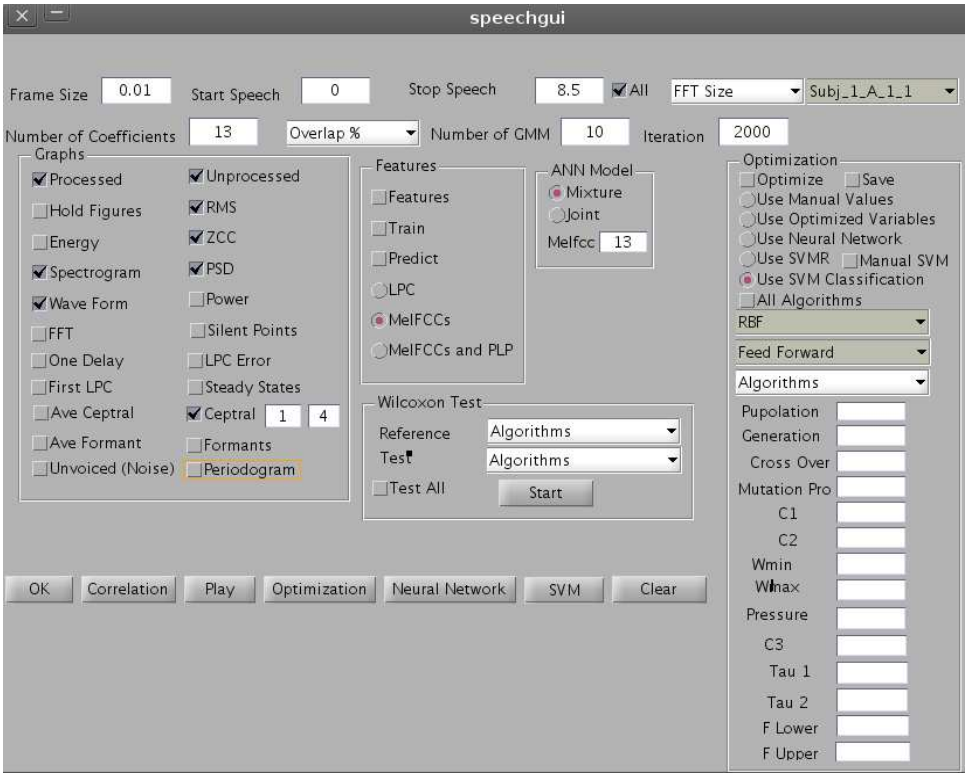

Figure 5.1: Main speech analysis and prediction tool

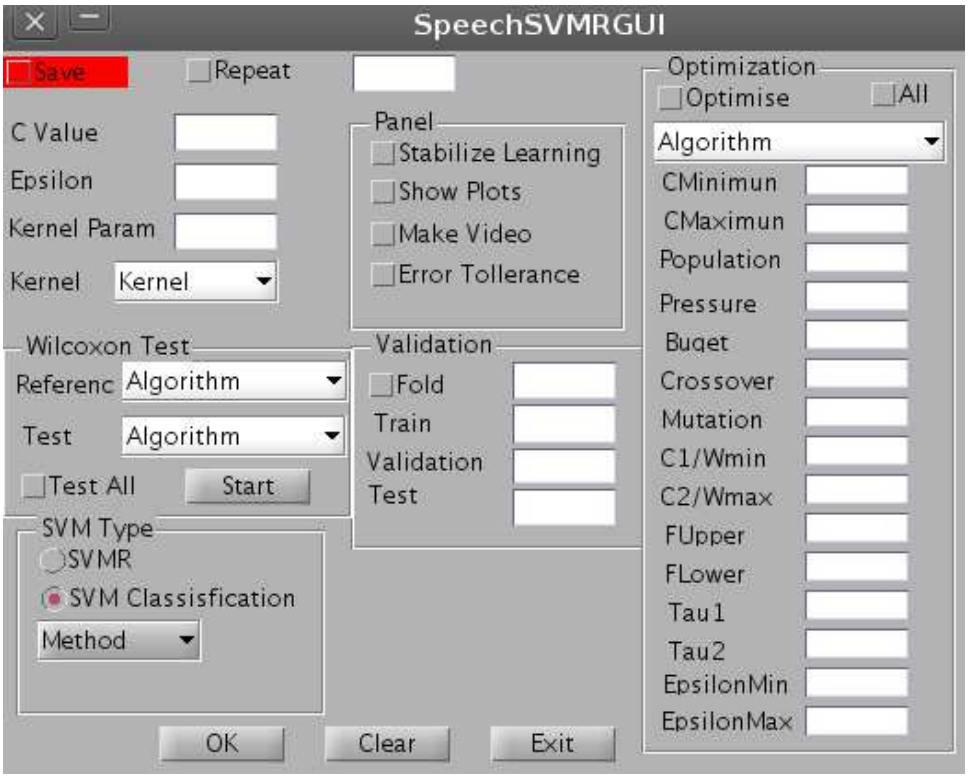

Figure 5.2: SVM training GUI

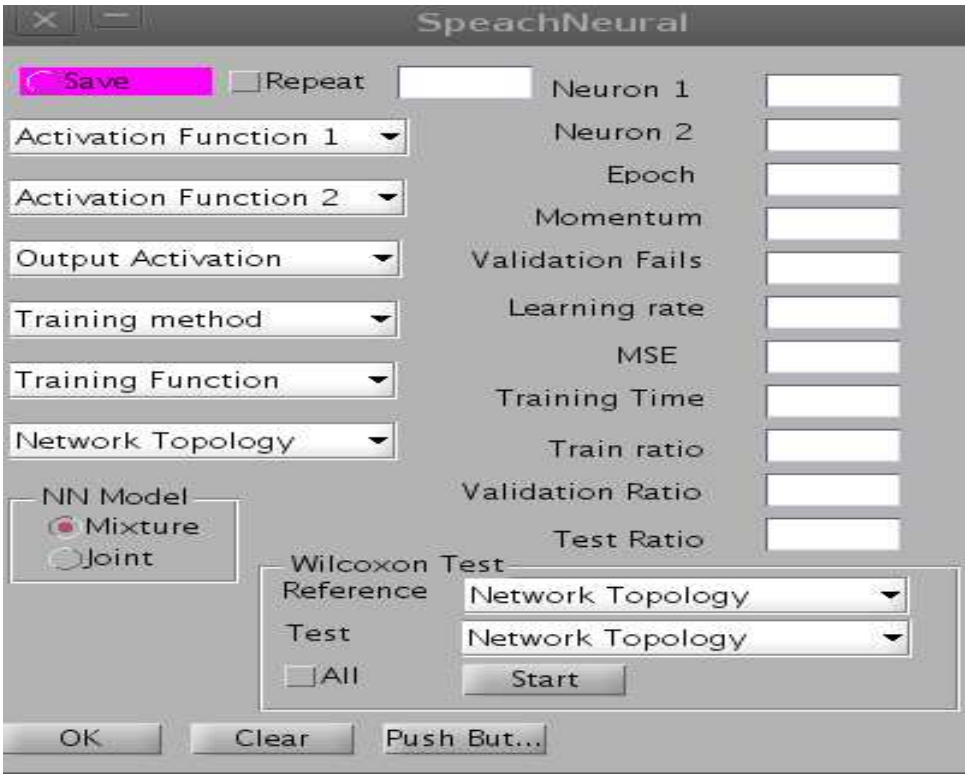

Figure 5.3: ANN training GUI

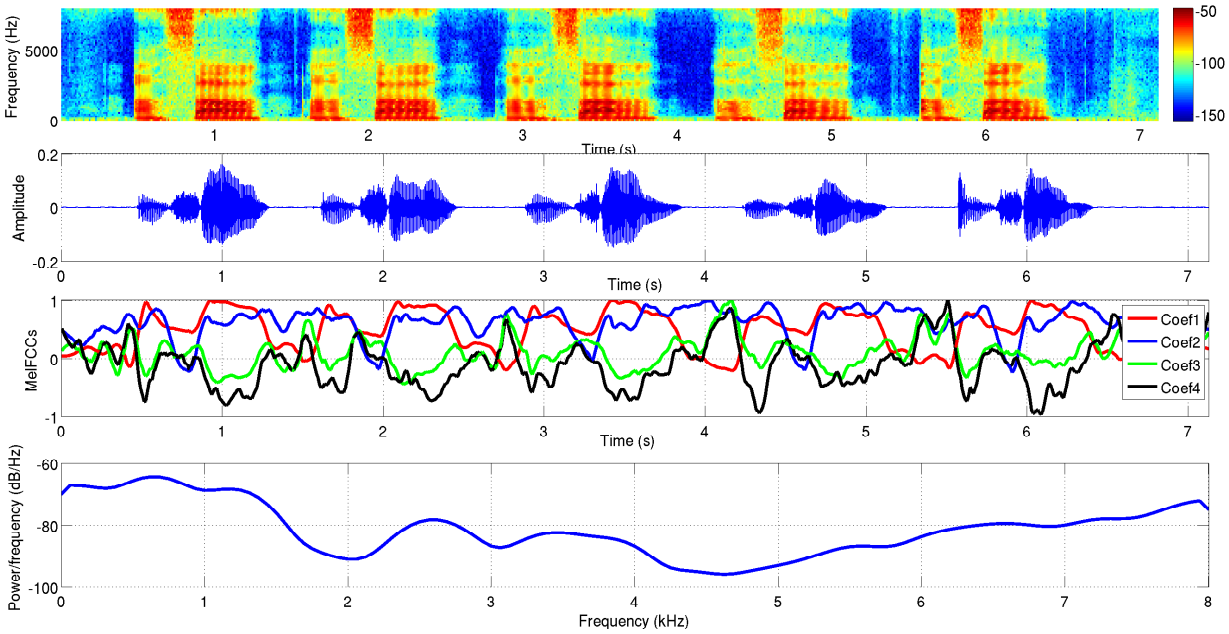

Figure 5.4: Speech spectrogram, wave, MelFCCs and PSD from male subject with silent/noise

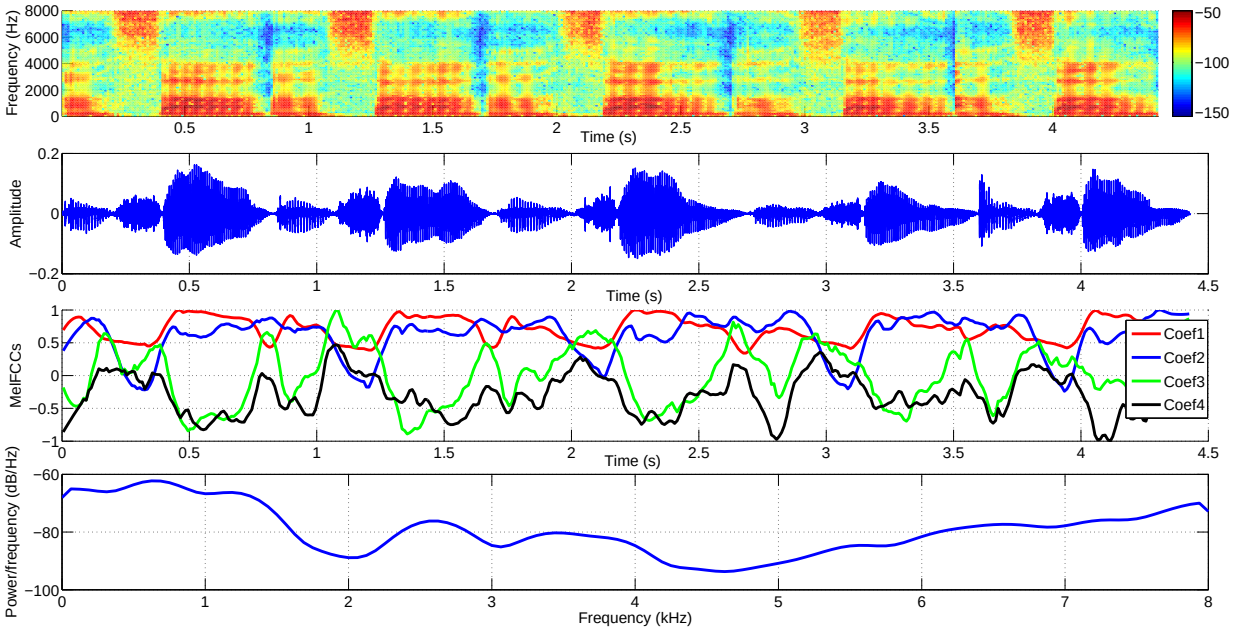

Figure 5.5: Speech spectrogram, wave, MelFCCs and PSD from male subject with silent/noise removed

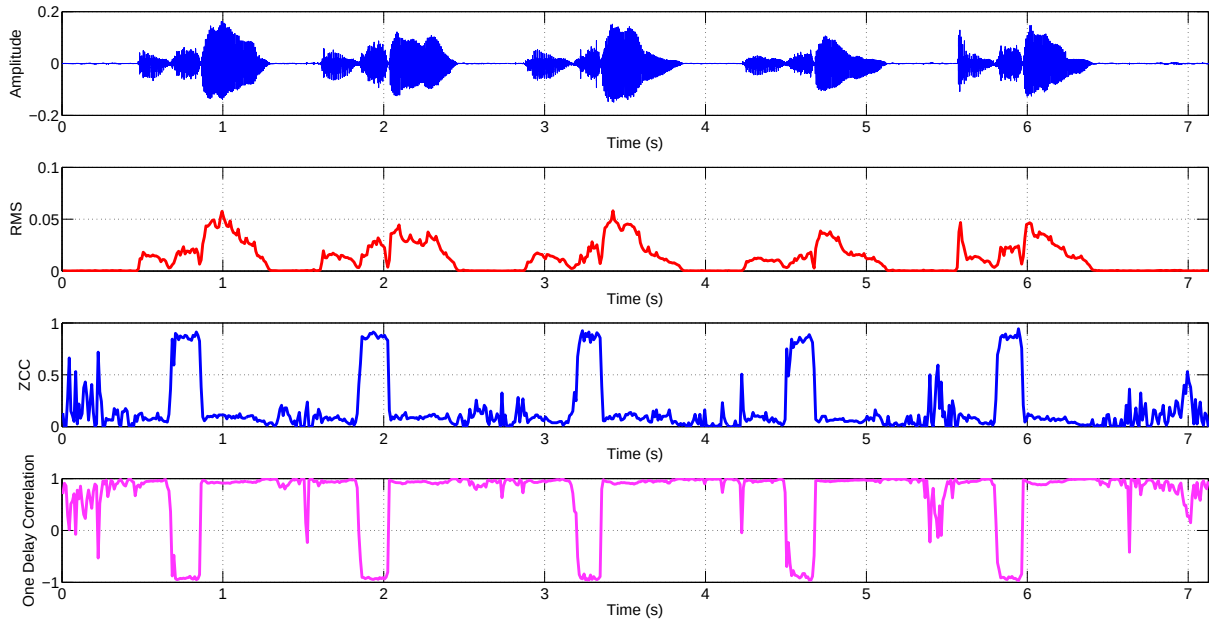

Figure 5.6: Speech wave, RMS, ZCCR and one delay autocorrelation coefficient from male subject with silent/noise

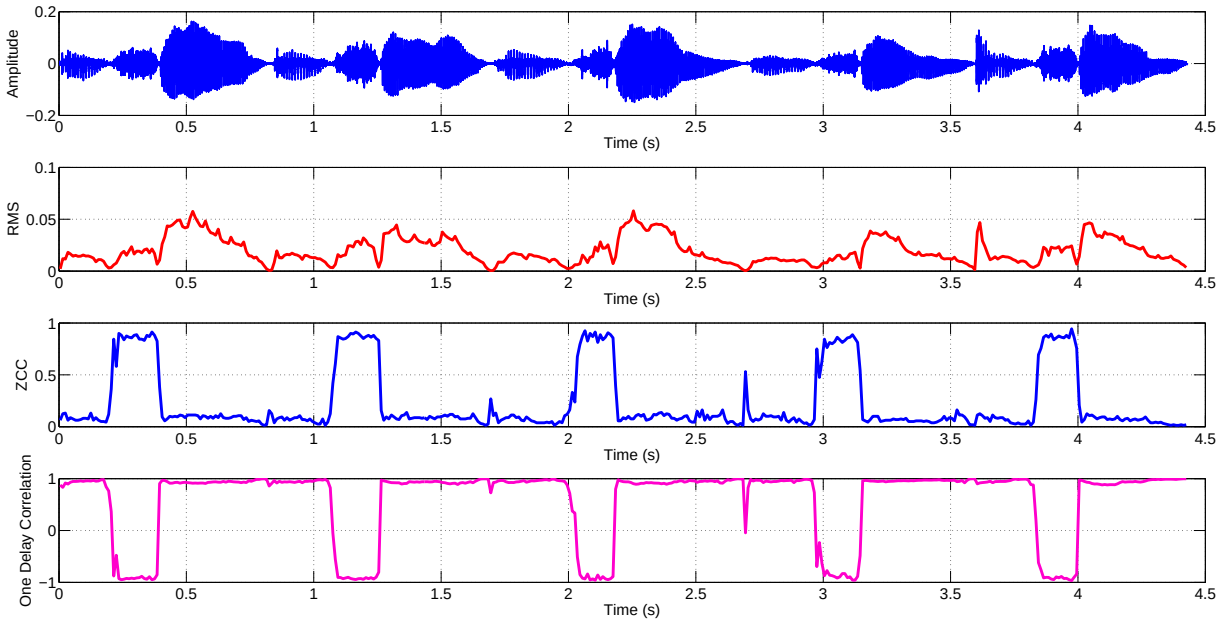

Figure 5.7: Speech wave, RMS, ZCCR and one delay autocorrelation coeficient from male subject with silent/noise removed

5.5 Artificial Neural Network

This section details the topologies of the ANN system used for the impaired speech segmentation. Regarding the general structure of this system, in order to address the nonlinearity associated with impaired speech pattern resulting from poor articulation, a fully connected multilayer perceptron (MLP) ANN with two hidden layers has been chosen Haykin (2008). The input layer is mapped into a high dimensional first hidden layer for proper features selection. After a preliminary study, the activation functions $F(x)$ is used for the two hidden layers which is a nonlinear hyperbolic tangent function Equation (5.12), and a linear symmetric straight line is used for the output activation function, Equation (5.13). The analytical expressions of the activation functions are:

$$F(x) = b \cdot tanh(ax) = b\left(\frac{1 - e^{-ax}}{1 + e^{-ax}}\right) \tag{5.12}$$

and

$$F(x) = mx + c \tag{5.13}$$

where the intercept $c = 0$ and the gradient m is left at Matlab default which is 1 while the constants a and b are assigned the value 1.

Since there is no existing algorithm or method of selecting the best ANN architecture or topology for a given non-linear problem, different architectures were adopted and their performances were compared. More specifically, four topologies were used i.e. feedforward (FF), feedforward with output feedback (FFB), cascaded feedforward (CFF), and layered recurrent feedback (LR).

5.5.1 ANN training

The dataset used in training the ANN was obtained from both male and female subjects. The total number of frames used for the training was 6456 obtain from nine repeated spoken disordered and target (correct) speech signals using a frame size of 10ms. All the data used in this research have been professionally recorded by the Interactive and Media Technologies research group at the Creative Technology Studios of De Montfort University, under the supervision of a speech therapist of the School of Applied Health Sciences at De Montfort University. The recording have been performed using the Avid HDX system.

The training is constrained to 300 epochs and 6 validation fails. The accuracy and level of generalization of ANN largely depend on the initial weights and biases, learning rate, momentum constant, training data, Gómez et al. (2014), Franco and Cannas (2000) and also the network topology. A supervised batch training method was used with 60% of the data used for training the ANN, 20% for validation and 20% for testing the trained ANN, David E. Rumelhart and Williams (1986). The learning rate is kept low at constant value of 0.01 and the momentum at 0.008. This choice has been performed since the backpropagation algorithm is used as a local search algorithm. In

this case, the ANN is used as a multiclass classifier, Subirats et al. (2010).

The inputs of the ANN for detection of silence/noise and unvoiced sections of the disordered speech are: the short time power (power), normalized autocorrelation coefficient at unit sample delay (NACUSD), the zero crossing count rate (ZCCR) and 13 MelFCCs as shown in Figure 5.8 and Figure 5.9 respectively. For steady state detection, in addition to the four inputs used for silence/noise and unvoiced speech detection, the normalized absolute value of the sum of the formants derivatives (SAVFD) is also used, as depicted in Figure 5.10.

The output of the ANN consist of thresholds which are use for decision making to classify a frame as either belonging to any of these three segments or not. Some of the voiced sections of the disordered speech does not belong to any of these three segments of interest, hence this study is different from normal speech recognition. During the training, if a frame belongs to segment i the target value of the ANN designed for identification of segment i will be assigned a random number in the range $[0.8, 1]$ otherwise it will be assigned a random number in the range $[0, 0.2]$, Equation (5.14) i.e. the output of the ANN are probabilistic. They are trained to give the probability of a given frame belonging to each of the segments. The target is the expected output of the ANN, the difference between the target and the actual output of the ANN during the training gives the error. The error is used to adjust the weights and biases of the ANN toward minimizing the mean square error (MSE) using backpropagation algorithm. Each of the three sections of the speech has it own ANN with a combine decision metric unit as shown in Figure 5.11. The term Artificial Neural Network Mixture Model (ANNMM) was used to reflect the fact that the three segments of the disordered speech considered in this study are detected using three different ANN topologies with a single decision unit as shown in Figure 5.11. When the trained ANN is tested with both normal and disordered speech utterances not used in training the ANN, the decision thresholds are different from the training thresholds i.e. $[0, 0.2]$ and $[0.8, 1]$ as detailed in Section 5.5.2. The choice of the upper limit 0.2 (if a frame does not belong to the reference (test) segment) and the lower limit 0.8 (if a frame belong to the test segment) create a gap of 0.6 between these two limits. This increases the robustness of the ANN especially when tested with speech not used in training the ANN as the decision thresholds will be set in-between these two limits, thus capturing a wide range of uncertainties as described in Section 5.5.2.

$$T_{(f,i)} = \begin{cases} rand(0.8,\ 1) & \text{if frame } f \text{ belong to segment } i \\ rand(0,\ 0.2) & \text{otherwise} \end{cases} \tag{5.14}$$

Where $T_{(f,i)}$ is the training target value (expected output of ANN during training) of frame f with respect to i, $i = 1$ refer to silence or noise segment, $i = 2$ unvoiced segment and $i = 3$ steady states segment.

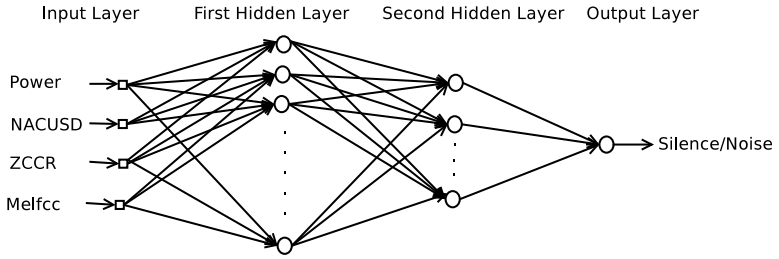

Figure 5.8: Feedforward ANN model for silence and noise detection

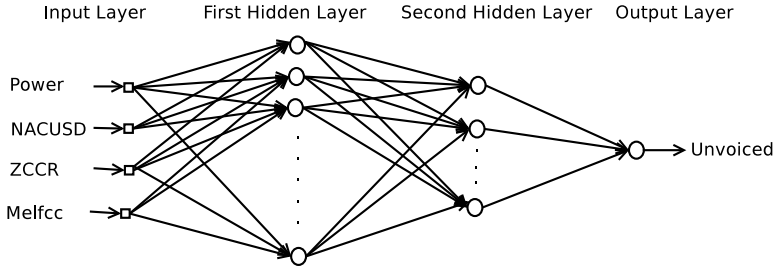

Figure 5.9: Feedforward ANN model for unvoiced detection

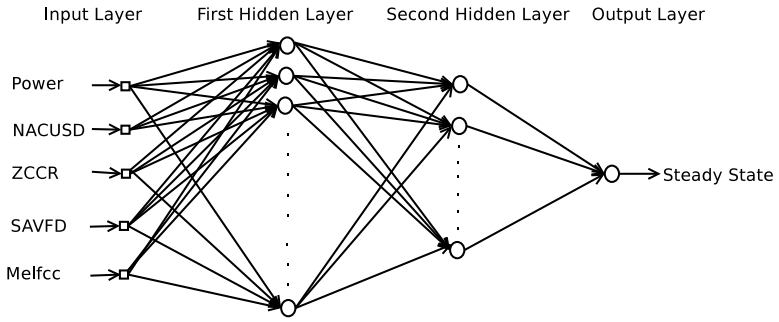

Figure 5.10: Feedforward ANN model for steady states detection

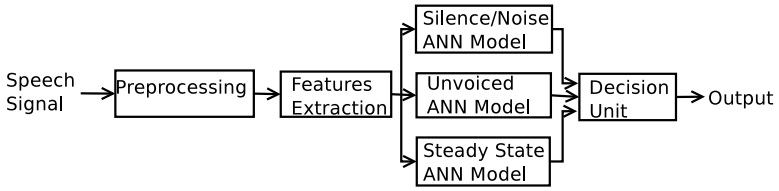

Figure 5.11: Artificial neural network mixture model (ANNMM)

5.5.2 ANN decision metrics

The three sections of the disordered speech of interest (silence/noise, unvoiced and steady state section) are mutually exclusive i.e. a frame can only belong to at most one of these three segments while some frames does not belong to any of the three segments. Out of the four ANN topologies implemented, each of the four is trained for detection of the three speech segments of interest. Three best ANN were chosen, one for identification of each segment. The ANN are trained to give the probability of a frame

belonging to each of the segments. Based on the training as described in Section 5.5.1, for an ideal ANN model, if a frame belongs to a given segment, the output of the ANN model for that segment should be in the range $[0.8, 1]$ otherwise it should be in the range $[0, 0.2]$ but this may not be the case especially when tested with impaired speech not used in training the ANN model. In order to increase the robustness of the ANN and to capture a wide range of uncertainties, the following empirical rule was set. A frame is said to belong to segment i if the output of ANN for segment i is $\geqslant \beta_i$ otherwise it does not belong to it, where β_i is the threshold for segment i. The decision thresholds β should range between 0.2 and 0.8 i.e. $0.2 < \beta < 0.8$ where 0.2 is the upper training bound if a frame do not belong to the particular segment while 0.8 is the lower training bound if a frame belong to the segment. After a manual tuning of the thresholds, the followings values: 0.6, 0.46 and 0.36 for silence/noise, unvoiced and steady state segment respectively are found to yield an excellent results that are robust and well generalized when tested with both impaired speeches used in training the ANNs and those not use in training the ANNs.

5.6 Support Vector Machine Model

SVMs have proven to be an effective tool for pattern recognition and binary classification on a wide array of problems Haykin (2008), Sun et al. (2013), Li et al. (2013). A SVM classifier essentially detects the hyperplane that allows to divide a set of objects (namely input set) into two classes. Since for the vast majority of the real-world classification problems, a dividing hyper-plane may not exist when low dimensional input features are used, hence a nonlinear mapping is applied to the objects in the input space. The co-domain of this mapping is called feature space.

More formally, the input space x is initially mapped into a high m dimensional feature space $\phi(x)$ by means of a nonlinear transformation. The mapping into a high dimensional feature spaces is very vital as many real world low dimensional features that a linearly non separable, often become linearly separable when transformed into a high dimensional features space, Xiaoyun et al. (2008). A linear model $f(\phi, w)$ Equation (5.15) is then constructed in the feature space, Haykin (2008), Sun et al. (2013), Li et al. (2013) which linearly mapped the feature space onto the output space.

$$f(\phi, w) = \sum_{j=1}^{m} w_j \phi_j + b \tag{5.15}$$

where $\phi_j = \phi_j(x),\ j = 1, \cdots, m$ refers to the set of non-linear transformations which transformed the input space into the feature space. The symbols w_j and b indicate the weights and bias, respectively, which transformed the feature space to the output space.

The segmentation of impaired speech considered in this study can be formulated as a multi-class classification problem. Each of the three classes of interest i.e. silent/noise, unvoiced and steady state sections are assigned a SVM binary classifier. In contrast to

the ANNs which are trained to give a probability of a frame belonging to each of the segments, the SVM are trained to give the exact class of a frame, i.e. the output of the SVM model designed for each segment can either be 1 if a frame belongs to that particular segment or -1 otherwise, as described in Equation (5.16).

$$T_{(f,i)} = \begin{cases} 1 & \text{if } f \text{ belongs to } i \\ -1 & \text{otherwise} \end{cases} \tag{5.16}$$

Where $T_{(f,i)}$ is the training target value of frame f with respect to i, $i = 1$ refer to silence or noise segment, $i = 2$ unvoiced segment and $i = 3$ steady states segment.

The SVM classification model is formulated as:

$$\text{minimize} \quad \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{n}\xi_i \tag{5.17}$$

$$\text{subject to} \quad y_i(w^T\phi(x_i) + b) \geqslant 1 - \xi_i. \tag{5.18}$$

The non-negative constant C is a regularization parameter that controls the trade-off between achieving low training error (misclassification) and minimization of the norms of the weights. The slack variable ξ is a parameter for handling non separable data points. The input vectors x_i for $i = 1, 2, \ldots$ are the support vectors while y_i for $i = 1, 2, \ldots$ are the corresponding outputs. The vector W is the vector of weights of the SVM. The parameter b is the bias of the SVM. The function $\phi()$ is the nonlinear transformation function that maps the input space to the feature space.

In this research, two kernel functions were used i.e. the Gaussian radial basis function (RBF) and the linear kernel as given in Equations (5.19) and (5.20) respectively. These two kernels are selected from the kernels used in Section 4.2.3 for designing SVM regression models for RF power prediction, but in this chapter, they are used for classification.

$$K(x, x_i) = e^{(-\gamma\|x - x_i\|^2)} \tag{5.19}$$

$$K(x, x_i) = x^T x_i + c, \tag{5.20}$$

where x indicates the input vector, γ is the kernel parameter. The adjustable constant parameter c of the linear kernel was set to 0.

The kernel function $K(x, x_i)$ is related to the linear transformation $\phi_j(x)$ as given by (5.21), Haykin (2008):

$$K(x, x_i) = \sum_{j=1}^{m}\phi_j(x)\phi_j(x_i). \tag{5.21}$$

The RBF kernel parameter γ and the regularization parameter C were evolved using metaheuristic algorithms, as will be detailed in Section 5.6.1, Mininno et al.

(2011a), Iacca et al. (2012b;a; 2011a;b), while the weights and bias of the SVM model have been set via convex optimization using Sequential Minimal Optimization (SMO) algorithm.

The accuracy and generalization of a SVM classifier depends on the correct choice of hyperparameter C for the linear kernel and the hyperparameters C and γ for the RBF kernel. Getting the optimum or near optimum values of the hyperparameters is further complicated as they are problem dependent. Figures 5.12, 5.13 and 5.14 shows the SVM models for the silent/noise, unvoiced and steady state segments respectively.

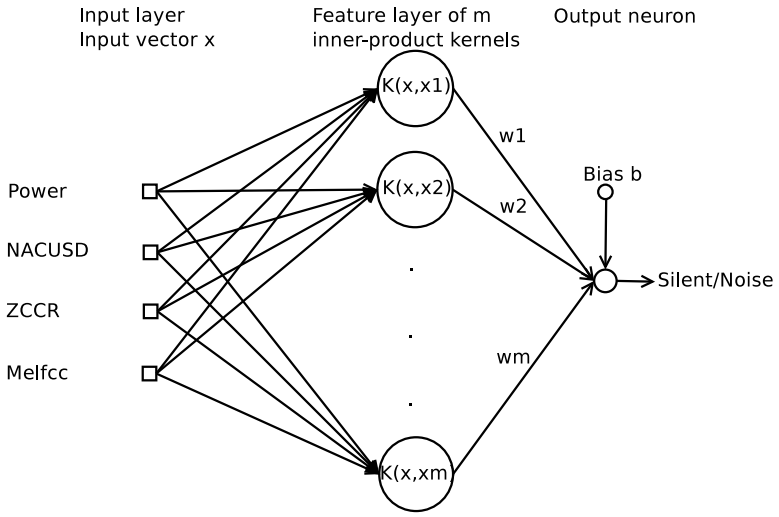

Figure 5.12: SVM model for silence and noise detection.

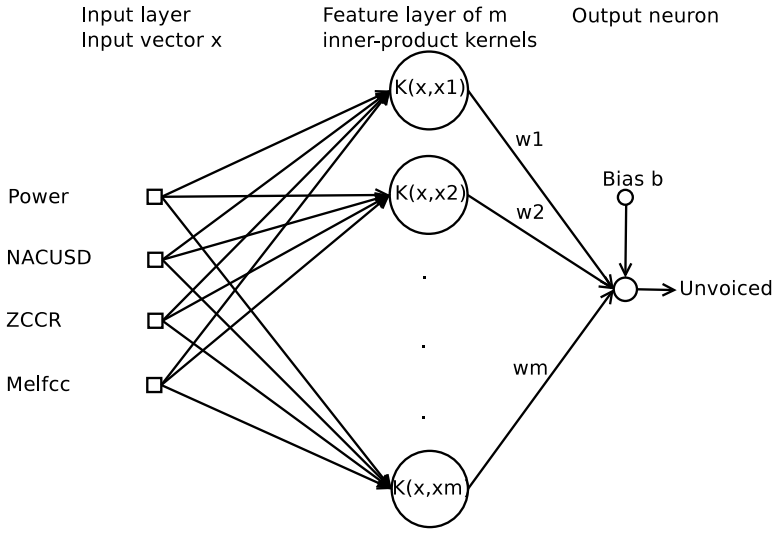

Figure 5.13: SVM model for unvoiced detection.

5.6.1 Optimization of SVM hyperparameters

The proposed nested algorithm used for training of SVM regression models for prediction of RF power discussed in Section 4.3.1, is extended for training of SVM classifier used

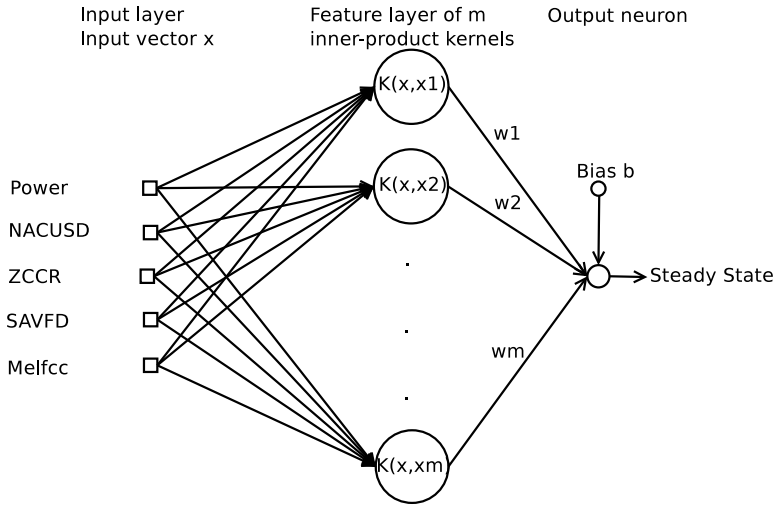

Figure 5.14: SVM model for steady state detection.

for the impaired speech segmentation. The training of the SVM occurs in two nested phases. In the outer phase, population based randomization metaheuristic optimization algorithms are used to sample the hyperparameters (i.e. the C parameter and the kernel parameter γ) within the decision space. Two variants of population based swarm intelligence optimization algorithms, and five variants of population based and compact differential evolution algorithms are used in the outer phase for sampling the C and γ. The sampled parameters are then used to construct an untrained SVM model. The inner phase consists of a supervised training method used to train the SVM model via convex optimization using sequential minimum optimization (SMO) algorithm to evolve the weights W and bias b of the SVM model. In order to estimate how the model will perform on speech utterances which were not used during the training phase, a ten folds cross validation is used during the training.

The differential evolution mutation schemes used for evolving the provisional offspring are sampled from the lists presented in Section 2.6.1 as follows:

- DE/rand/1:

$$V_{i,G} = X_{r_1,G} + F \cdot (X_{r_2,G} - X_{r_3,G}) \tag{5.22}$$

- DE/best/1:

$$V_{i,G} = X_{best,G} + F \cdot (X_{r_1,G} - X_{r_2,G}) \tag{5.23}$$

- DE/rand-to-best/1:

$$V_{i,G} = X_{i,G} + F \cdot (X_{best,G} - X_{i,G}) + F \cdot (X_{r_1,G} - X_{r_2,G}) \tag{5.24}$$

- DE/best/2

$$V_{i,G} = X_{best,G} + F \cdot (X_{r_1,G} - X_{r_2,G}) + F \cdot (X_{r_3,G} - X_{r_4,G}) \tag{5.25}$$

- DE/rand/2

$$V_{i,G} = X_{r_5,G} + F \cdot (X_{r_1,G} - X_{r_2,G}) + F \cdot (X_{r_3,G} - X_{r_4,G}) \tag{5.26}$$

Where the indexes r_1, r_2, r_3, r_4 and r_5 are mutually exclusive positive integers and distinct from i within the range $[1 \quad PN]$. $X_{best,G}$ is the individual with the best fitness at generation G while F is the mutation constant or scaling factor. PN is the population, X are the parent i.e. potential or possible solutions and V are the mutants or provisional offspring.

The mutant vector $V_{i,G}$ is then combined with $X_{i,G}$ by means of the crossover operation to generate the offspring as discussed in Section 2.6.1. Two types of crossover are used, namely binomial and exponential crossover, Zaharie (2007). The fitness function used is the percentage error (PE) of miss classification or detection, given by Equation (5.31). A greedy selection scheme is used as described in Section 2.6.1, i.e. if the offspring outperformed its parent, the offspring will replaced its parent in the next generation otherwise the parent will continue in the next generation. The smaller the fitness function PE the beter the performance. Population based differential evolution (DE) and it compact (cDE) version were detailed in Section 2.6.1, Zaharie (2007), Price et al. (2005), Qin et al. (2009), Mininno et al. (2011a), Iacca et al. (2012b), Mininno et al. (2011a), Iacca et al. (2012b).

The two variants of swarm intelligence algorithms used are inertia weighted particle swarm optimization (PSO) and complementary learning particle swarm optimization (CLPSO) algorithms, Shi and Eberhart (1998), Liang et al. (2006) as discussed in Section 2.6.5. The velocity update of PSO and CLPSO are giving by Equation (5.27) and (5.28) respectively while the particle update for both PSO and CLPSO is shown in Equation (5.29). The learning probability P_{ci} of CLPSO is given by Equation (5.30).

$$\begin{aligned} V_i^d \quad \longleftarrow \quad & w_i \cdot V_i^d + C_1 \cdot rand1_i^d \cdot (pbest_i^d - X_i^d) \\ & + C_2 \cdot rand2_i^d \cdot (gbest_i^d - X_i^d) \end{aligned} \tag{5.27}$$

$$V_i^d \longleftarrow w_i V_i^d + c \cdot rand_i^d \cdot \left(pbest_{f_i(d)}^d - X_i^d\right) \tag{5.28}$$

Where *pbest* is particle best, *gbest* is the global best, $pbest_{f_i(d)}^d$ is any particle's *pbest* including particle i *pbest*. $f_i = [f_i(1), f_i(2) \cdots, f_i(D)]$ defined which particles' *pbests* the particle i should follow. D is particle dimension, $rand_i^d$ is a random number in the range [0 1], each particle dimension d has its own $rand_i^d$. X referred to particles' positions while c is the acceleration pull. PSO has been discussed extensively in Section 2.6.5.

$$X_i^d \longleftarrow X_i^d + V_i^d \tag{5.29}$$

$$P_{ci} = 0.05 + 0.45 \frac{e^{\frac{10(i-1)}{SS-1}} - 1}{(e^{10} - 1)} \tag{5.30}$$

where SS is the swarm size (number of particles) and P_{ci} is the learning probability .

5.7 Numerical Results

The results related to the two neural models are divided into the following two subsections

5.7.1 ANN Results

After training the ANN, the trained ANN model was tested with impaired speech signals from both the training data and those not used in training the ANN model. These test speeches are derived from both male and female subjects. The results obtained demonstrate the robustness of the approach and how well the network is generalized.

ANN topology and parameters setting are problem dependent, the number of neurons in the first and second hidden layers, and the output layers are summarized in Table 5.1. The number of neurons in the layered recurrent feedback ANN, is less when compared with other topologies as presented in Table 5.1 but the number of parameters (weights) to be optimized is very large because of the layers feedback loops. This results in a long training time. The layered recurrent feedback topology is run 15 times constrained within 100 epochs and 6 validation fails because of it long training time while the other topologies were run 30 times constrained within 300 epochs and 6 validation fails. For each run the ANN weights and biases were reinitialized with random values for new training.

Table 5.1: ANN Topologies Specification

	FF	FFB	LR	CFF
1 Hidden Neurons	30	30	15	30
2 Hidden Neurons	20	20	5	20
Output Neurons	1	1	1	1

In order to form the artificial neural network mixture model (ANNMM), the best ANN architecture or topology for each of the three segments of the impaired speech (i.e. silent/noise, unvoiced and steady state segment or section) were selected. These best topologies were combined together with a single decision unit to form the ANNMM, Figure 5.11. To evaluate the performance of each ANN topology, the percentage error (PE) was used, Equation (5.31). The PE is calculated for each ANN architecture and for each segment of the speech.

$$PE_i = \frac{M_i + Fa_i}{N_f} \times 100 \tag{5.31}$$

where PE_i is the percentage error of segment i; for $i = 1$ silence/noise, $i = 2$ unvoiced

Table 5.2: Test Results Statistcs

ANN	Silence/Noise		Unvoiced		Steady State	
Topology	PE	STD	PE	STD	PE	STD
FF	1.63	0.34	1.60	0.37	2.84	0.50
FFB	1.87	0.36	1.50	0.38	2.32	0.32
CFF	**1.57**	**0.18**	1.89	0.18	2.42	0.29
LR	1.88	0.13	**1.48**	**0.19**	**2.20**	**0.36**

Table 5.3: Comparative Results (PE)

ANN	Silence/Noise	Unvoiced	Steady State
proposed	**1.57**	**1.48**	**2.20**
Iliya et al. (2014c)	2.29	3.44	5.93

and for $i = 3$ steady state. M_i is the total number of frames belonging to i which were not detected (miss detection) i.e. classified as not belonging to i, Fa_i is the number of frames which do not belong to i but were classified as belonging to i (false alarm). N_f is the total number of frames. Hence, the PE consist of the sum of miss detection and false alarm. The PEs and the corresponding standard deviations (STD) for each topology are listed in Table 5.2 where the best results are highlighted in bold font. All the results contained in the tables in this section are obtained by using the test dataset i.e. data not used for training the ANN model.

For this problem, the layered recurrent feedback (LR) rank first for unvoiced detection followed by feed-forward with output feedback (FFB). For steady state detection, layered recurrent feed-back is also the best followed by feed-forward with output feedback (FFB). For steady state detection, feed forward is the best followed by LR as shown in Table 5.2. For silence/noise detection, cascaded feed forward (CFF) is the best followed by feed forward (FF) model. It is interesting to note that the recurrent topologies outperform the FF in must cases, this may be due to the fact that the impaired speech is fed into the ANN as a stream of data in time order just the way the patient pronounces the words (vowel-consonant-vowel) hence there is a likelihood of a structure or relationships that exist between consecutive frames which is captured in the feedback sections of the recurrent ANN models.

In order to highlight the performance of the proposed ANN, the results obtained have been compared with those achieved by the neural system published in Iliya et al. (2014c) for solving the same task. Numerical results are displayed in Table 5.3.

Results in Table 5.3 clearly show that the newly proposed ANN performs better than the study by Iliya et al. (2014c). More specifically, it leads to an improvement of 31.44%, 56.98%, and 62.90% for silence/noise, unvoiced and steady state, respectively.

It must be noted that a direct comparison with the work carried out by Iliya

et al. (2014d) is not applicable, since in Iliya et al. (2014d) it was assumed that at the beginning of the speech recording, the first ten frames (100 ms) are either noise or silence (in other words the problem was slightly different). This assumption was necessary to obtain the initial thresholds for silence/noise. Although such assumption is often used in speech related applications, it is fully arbitrary and may not always be true. In the present study, such assumption was not done and, thus, the proposed detection model is consequently more robust.

While the results presented in Table 5.2 are obtained using the test dataset which even though it is not use in training the ANN model, it is a subset of the data from which the training and validation datasets are obtained. In order to evaluate the robustness of the ANNMM, the model was tested using an independent impaired speech obtained from male and female subjects. Figures 5.15 and 5.16 show the segmentation of two independent impaired test speeches from a male and female subjects respectively using the trained ANN model. In both Figures 5.15 and 5.16, the two top subplots show the speech signal (wave) and result of ANN segmentation. The top subplot highlights the silence/noise segments between red beams captured by the ANN, the middle subplot highlights unvoiced and steady state segments between black and red beams respectively. Finally the bottom subplot displays the steady state segment of the signal using the normalized formant derivative, i.e. the normalized trajectory of the average of the absolute rate of change (first derivative) of the first three formants obtained from Equation (5.10) (red beams).

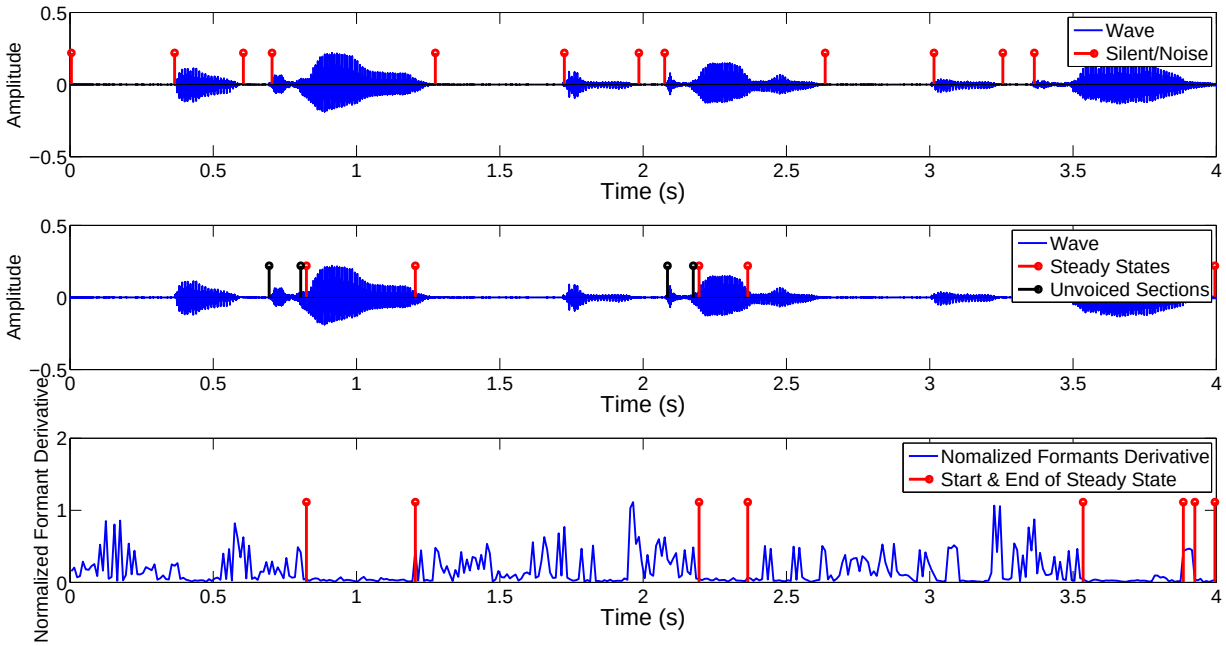

Figure 5.15: Test speech utterance from male using ANN: speech signal with segmentation and normalized formant derivative

From the first subplot of Figures 5.15 and 5.16, it can be seen that the ANN

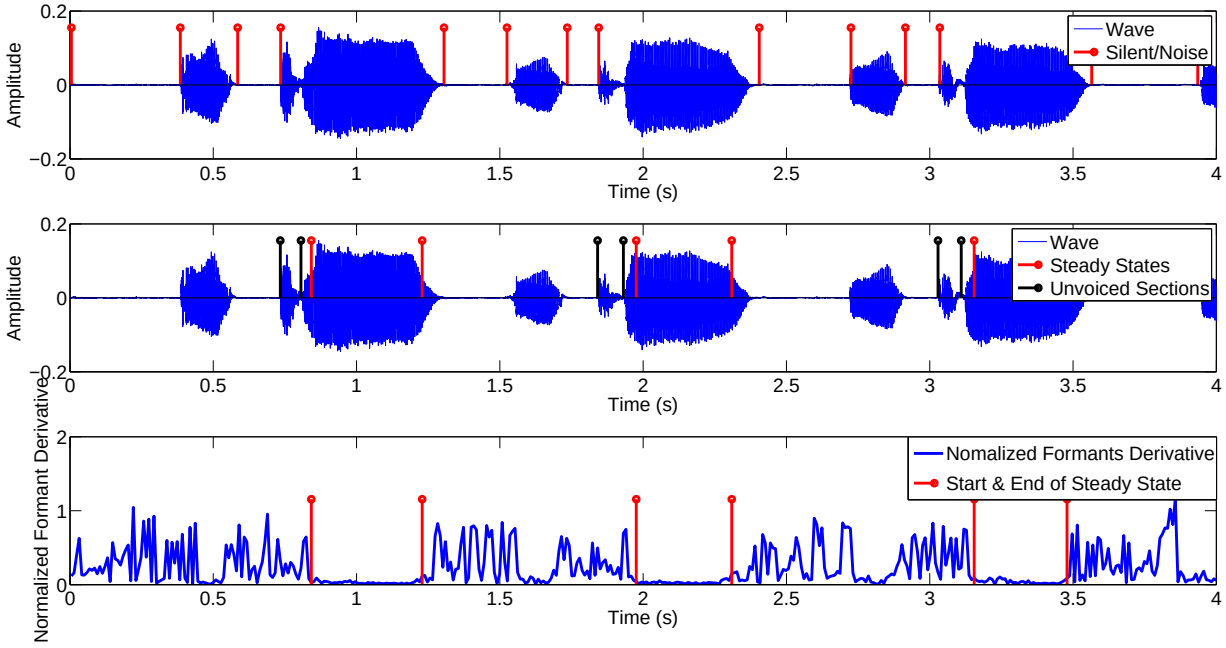

Figure 5.16: Test speech utterance from female using ANN: speech signal with segmentation and normalized formant derivative

was able to detect both silence and noise sections efficiently (red beams). The second subplot shows how the ANN was able to detect the beginning and end of unvoiced (black beams) and the steady state (red beams) sections. The informative section of the impaired speech lies between the beginning of the unvoiced section to the start of the steady state. The third subplot clearly shows that within the steady state, the formant remains fairly constant and consequently the derivative is approximately null. This is a very important finding for creation of an artificial speech therapist since the portion of speech signal with no new spectral information are identified and captured, thereby restricting the area of interest of the signal.

5.7.2 SVM Results

In order to select the most appropriate optimization algorithm for the SVM problem, multiple algorithms have been considered as described in Section 5.6.1. The following popular population-based and compact metaheuristics have been implemented and run: PSO, CLPSO, Liang et al. (2006), jDE Brest et al. (2006) and DE, Neri and Tirronen (2010). Then, to prove the effectiveness and the potential of cDE for the specific problem, multiple cDE configurations have been tested. All DE and cDE algorithms metaparameters are set after a manual tuning as follow: the crossover rate is tuned to 0.3 and scale factor also set to 0.3. The PSO has been run with velocity and acceleration constants set to $C_1 = 0.8$ and $C_2 = 0.6$ respectively. The CLPSO and jDE algorithms have been run with their original setting suggested by Liang et al. (2006) and Brest et al. (2006) respectively. The population based algorithms have been run

Table 5.4: SVM Test Results Statistcs

Algorithms	Linear Kernel						RBF Kernel					
	Silent/Noise		Unvoiced		Steady State		Silent/Noise		Unvoiced		Steady State	
	PE	STD	PE	STD	PE	STD	PE	STD	PE	STD	PE	STD
PSO	1.67	0.00=	2.39	0.02=	8.91	0.00+	1.28	0.00=	1.56	0.02+	2.96	0.04+
CLPSO	1.67	0.00=	2.41	0.02+	8.94	0.04+	1.60	0.20+	1.80	0.28+	4.45	2.41+
DE/rand/1/bin	**1.67**	**0.00**	**2.39**	**0.02**	8.91	0.00+	**1.28**	**0.00**	1.53	0.02+	2.90	0.02+
DE/rand/1/exp	1.67	0.00=	2.40	0.00+	8.91	0.00+	1.28	0.00-	1.56	0.02+	2.93	0.03+
JDE	1.67	0.00=	2.67	0.00+	9.33	0.00+	1.84	0.08+	**1.24**	**0.00**	2.25	0.00+
JDE/exp	1.67	0.00=	2.67	0.00+	9.33	0.00+	1.89	0.04+	1.29	0.07+	2.56	0.14+
DE/best/2/bin	1.67	0.00=	2.67	0.00+	9.34	0.02+	1.91	0.04+	1.38	0.16+	2.25	0.00+
cDE/rand/1/bin	2.01	0.00+	3.02	0.00+	8.37	0.00+	1.86	0.05+	1.39	0.09+	3.81	0.14+
cDE/rand/1/exp	2.01	0.02+	3.02	0.00+	8.35	0.03+	1.83	0.03+	1.40	0.07+	3.81	0.11+
cDE/best/2/bin	1.94	0.04+	3.06	0.00+	8.34	0.03+	1.35	0.02+	1.55	0.46+	**1.95**	**0.03**
cDE/best/2/exp	2.00	0.03+	3.04	0.02+	8.35	0.03+	1.35	0.02+	1.47	0.41+	2.01	0.04+
cDE/best/1/bin	1.99	0.05+	3.05	0.02+	**8.33**	**0.02**	1.37	0.05+	1.46	0.42-	2.28	0.70+
cDE/best/1/exp	1.98	0.04+	3.05	0.02+	8.34	0.03+	1.37	0.04+	1.55	0.44+	2.00	0.11+
cDE/rand-to-best/1/bin	2.01	0.00+	3.04	0.02+	8.37	0.00+	1.34	0.00+	1.27	0.03+	2.28	0.15+
cDE/rand-to-best/1/exp	2.01	0.00+	3.04	0.02+	8.37	0.00+	1.34	0.00+	1.29	0.02+	2.43	0.37+
cDE/rand/2/bin	2.00	0.02+	3.02	0.00+	8.36	0.02+	1.82	0.00+	1.34	0.10+	2.11	0.18+
cDE/rand/2/exp	2.01	0.00+	3.02	0.00+	8.37	0.00+	1.86	0.05+	1.36	0.12+	2.11	0.18+

with a population size of 10 individuals while the virtual population of cDE variants has also been set to 10, Mininno et al. (2011a).

Each of the algorithms involved in this study has been run with 200 fitness evaluations. For each fitness evaluation, the weights and bias were evolved using Sequential Minimal Optimization (SMO) algorithm constrained within a maximum of 70000 iterations. Each algorithm has been run for 50 independent runs. The PE of misclassification and its corresponding standard deviation (STD) using test impaired speech are listed in Table 5.4. Furthermore, in order to enhance the statistical significance of the results, Wilcoxon test has been applied to select the most promising optimizer, Wilcoxon (1945). The most promising metaheuristic for each column in Table 5.4 has been highlighted in bold and taken as a reference for the Wilcoxon test. A + indicate that the reference algorithm outperformed the competing algorithm, − signified that the competing algorithm outperform the reference while = mean that the competing and reference algorithms are statistically equal in performance.

From the results depicted in Table 5.4, the RBF kernel generally outperformed the linear kernel in capturing the three segments of the impaired speech considered in this study. The RBF SVM trained using the proposed algorithm also outperforms the ANN based segmentation model for the three objectives, corresponding to PEs of 1.28%, 1.24%, and 1.95% while the corresponding PEs of the ANN system are 1.47%, 1.48%, and 2.20% for silent/noise, unvoiced and steady state respectively.

The comparison among the optimization algorithms shows that DE based algorithms appear to have a slight advantage with respect to the PSO metaheuristcs considered in this research. While population based algorithms appear to display a slightly better performance than the compact DE for the classical tasks of silent/noise and unvoiced detection. On the contrary, cDE appears to have the best performance for the specific steady state detection which is the most crucial task in this study. In conclusion, the cDE framework appears to have a very good performance for this task in both linear and RBF kernel. Thus for this system, a population based DE is preferred for noise/silence and unvoiced detection while its compact version for steady state detec-

Table 5.5: SVM global best model parameters

	Linear Kernel			RBF Kernel		
	Silent	Unvoiced	Steady State	Silent	Unvoiced	Steady State
PE	1.67	2.36	8.29	1.28	1.16	1.90
C	5.09	4.29	20.94	530.01	58.26	204.69
γ	-	-	-	8.82	1.88	1.66
	DE/rand/1/bin	DE/rand/1/bin	cDE/best/1/bin	DE/rand/1/bin	jDE	cDE/best/2/bin

tion.

Figure 5.17 shows the detection of the three segments using SVM on a male subject speech. This speech is not a subset of the dataset from which the training, validation and test dataset were obtained, rather is completely an independent impaired speech. Considering how well the segmentation was carried out with reference to Figure 5.17 as well as other figures presented in this study, the model appears to be robust and well generalized.

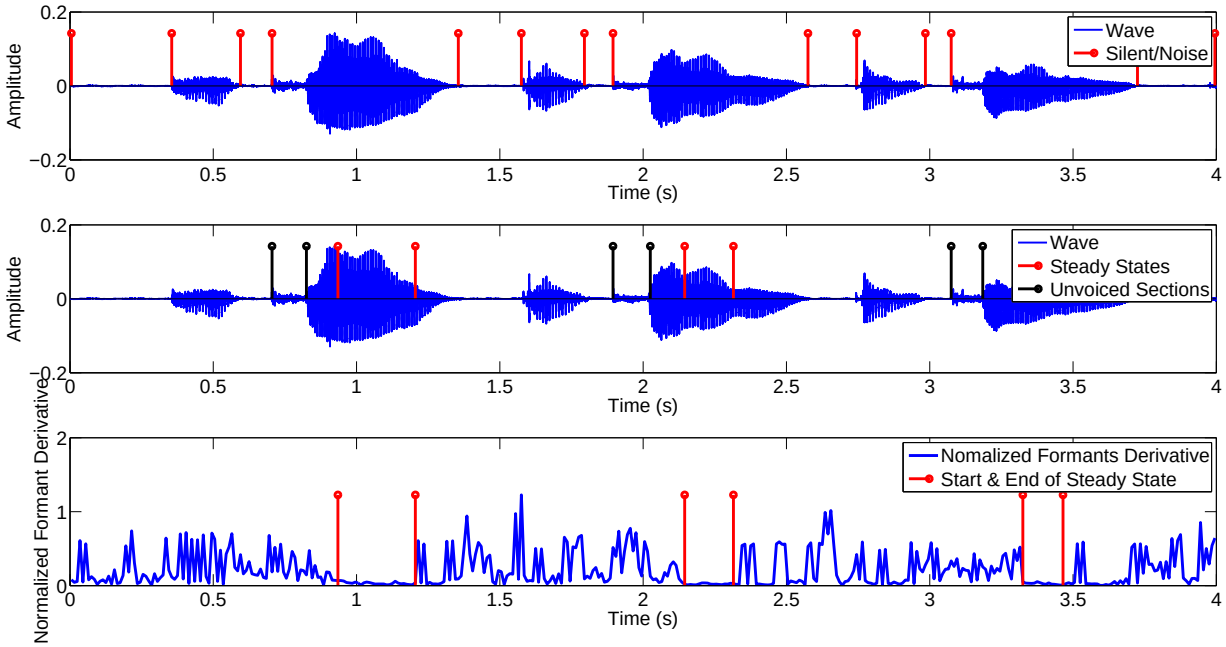

Figure 5.17: Test speech utterance from male using SVM: speech signal with segmentation and normalized formant derivative

The best results found during the experiments are reported in Table 5.5 with the corresponding best values of C and γ.

5.8 Summary of Impaired Speech Segmentation

The results obtained demonstrate that ANN and SVM can be used for effective segmentation of disordered speech utterances. A robust segmentation model was developed by combining three different ANN topologies to form the artificial neural network mixture model (ANNMM) which efficiently isolates the silence/noise and unvoiced sections of the impaired speech and identifies where the spectrum becomes steady. Numerical results show that a very high performance can be obtained using a SVM with RBF kernel trained by the proposed nested algorithmic framework. Population based DE algorithms display the best performance for detecting noise/silence and unvoiced segments while the compact DE algorithm offers the best performance for the steady state detection. Thus, a hybrid model, making use of both population based and compact algorithms seems to be the most promising option. Having been able to capture the most informative part of the impaired speech, future work will focus on analysis of the informative sections of the impaired speech, and to identify in time the articulators that are responsible for each sections of the impaired speech and possible correction measures. This will be done using other high dimensional temporal (time) and frequency domain speech features along with other CI models. The final phase of this research will focus on the development of a robust artificial speech therapist that can serve as a self training system for speech disordered patients with reliable feedback progress report in form of text and audiovisuals.

Chapter 6

Conclusion and Future Work

6.1 Summary of publications

During the period of this research I had five peer-reviewed journal papers published, and presented my work at six peer-reviewed IEEE conferences. The sixth journal was submitted to IEEE transactions on neural networks and learning systems for possible publication.

6.2 Contribution of the thesis

The contributions of this study are listed as follows:

- Spectrum occupancy or utilization estimate of some communication bands in Leicester, UK was conducted. These bands include: 50MHz to 860MHz consisting of radio and TV broadcasting channels; and 868MHz to 2.5GHz with emphases on ISM 868 band, GSM900 and GSM1800 bands and ISM 2.4GHz band, some bands within these ranges where also considered. The success of CR network is largely dependent on the overall world knowledge of spectrum utilization in both time, frequency and spatial domain, thus this study will add to the existing spectrum survey campaigns conducted in other part of the world in order to facilitate and enhance the implementation of CR network. This survey aim at identifying underutilised bands that can serve as potential candidate bands to be exploited for CR application. It is quite interesting to note that the outcome of the survey reveals that the average utilization estimate of the bands within the studied location is less than 30% of the entire bands considered, Iliya et al. (2015c).

- Optimised ANN and SVR models for prediction of RF power traffic of the studied location were developed. These models are used for prediction of real world RF power and instantaneous occupancy status of the channels. While most predictive models implemented were centred on prediction of occupancy status (i.e. either a channel is busy or idle) using simulated data, the proposed models predict real

world RF power from which the occupancy status can also be known for a given noise floor (threshold). For vacant (idle) channels the predicted power referred to the expected noise power while for busy channels is the communication signal plus noise and/or interference power. The prediction models presented will enable the selection of channels with less noise among the predicted idle channels for effective communication at less radio resources such as bandwidth and transmitted power. More importantly, the inputs of the proposed prediction models does not contain any RF power related parameters such as signal to noise ratio or a priory knowledge of primary users' signals which may not be made available to CRs. This make the proposed models to be robust and well suited for CR application especially where the CR has no a priory knowledge of the PU signal type, Iliya et al. (2015d; 2014a;b).

- The thesis proposes a robust and novel way of obtaining a bench mark for the weights of a combined prediction model that captures the advantages of equal weight average method, inverse ranking method and inverse mean square error method of evolving weights of a combined forecasting model as detailed in Section 4.4.1. The proposed hybrid bench mark tuned using differential evolution variants, proved to converge quickly to a good solution as shown in Figure 4.10. Six prediction models consisting of different topologies of ANN and SVR models are combined together with weighted outputs to form a single prediction model. As a prove of concept, the weights of the combined model were evolved using the proposed method and the overall prediction accuracy was found to increase with an average of 46% as compare with the individual models used.

- The work presented by Cherkassky and Ma (2004) was improved by using differential evolutionary and particle swarm optimization algorithms to evolve the meta parameters of the SVR models which was found to increase the accuracy and generalization of the models as described in Section 2.7. The SVR models was trained by the proposed nested ad-hoc hybrid algorithm whose outer phase consist of metaheuristics optimization algorithm variants for sampling the metaparameters while the inner phase is designed using convex optimization algorithm for evolving the weights of the SVM regression model, Section 4.3.1. The initial weights of the ANN are also evolved using deferential evolutionary and particle swarm optimization algorithms. These population based randomization algorithms are used as global searchers which provides effective means of escaping local optimum during the initial training phase of the ANN, and finally the best solution (weights) are then fine tuned toward the global optimum using single solution back propagation algorithms (BPA) at low learning rate and momentum constant to serve as a local searcher. The ANN models trained with the hybrid algorithm was found to outperform the ones train with only BPA with a MSE of 0.0242 while BPA have a MSE of 0.1345, Iliya et al. (2015a;d; 2014b).

- The skills and knowledge acquired from this research were extended to develop computational intelligence models for application in three areas related to medical

field. CI models for segmenting impaired speech into silence/noise, unvoiced and voiced sections, and for identification of the segment where the spectrum becomes steady were developed. Thus capturing the part of the speech that is likely to contain the information about the speech articulators (organs) responsible for the impairment, Iliya and Neri (2016), Iliya et al. (2014d;c).

Chapter 2 forms one of the subsidiary contribution of this research where intensive literature review of the areas within the research field are presented. This led to the discovery of research gaps and possible solutions were suggested. The background information needed for understanding of the subsequent chapters of the thesis are presented. Spectrum sensing techniques are discussed with emphases on energy detector which is the technique used in sensing the channels considered in this research, and for estimating the utilization rate of the channels. The ANN topologies used to design the prediction models are discussed in this chapter. A supervised training was used and after several experimental trial, the best activation functions for the hidden layers and the output layer are found to be tannsig and symmetric straight line respectively. The theory of the metaheuristics optimization algorithms used for optimizing the RF power and spectrum hole prediction models are discussed. Different variants of population based and compact versions of differential evolutionary and swarm intelligence optimization algorithms are used. The chapter present a novel way of improving the work done by Cherkassky and Ma (2004) via the use of the aforementioned metaheuristics optimization algorithms for evolving the meta-parameters of SVM as detailed in Section 2.7. This approach was found to improve the accuracy and generalization of the prediction models as evidenced by the results presented in Section 4.5.1.

Chapter 3 forms one of the major contribution of this thesis where Long time spectrum occupancy survey using real world RF data captured at different period of time within three years was conducted. The survey aims at evaluating the utilization levels of the various bands under study, and to identify underutilized bands or channels that can serve as potential candidate channels to be used for CR application towards efficient spectrum utilization. The survey reveals that, for a given probability of false alarm of 0.015, the average occupancies of bands 50 - 2500 MHz at constant noise margins of 5, 7 and 10 dB above the ambient noise are 31.52%, 20.86% and 11.70% respectively, Figure 3.17. When the decision threshold was obtained by setting the constant noise margin at 10 dB above the average of the ambient noise as recommended by International Telecommunication Union (ITU), Bureau (2002), the estimated average usage of the spectrum was found to be 18.34%. This implies that out of the 1182 MHz total bandwidth considered, the overage spectrum often in used is 216.78 MHz which is grossly inefficient. Band 50-860 MHz assigned for different services as described in Section 3.2 is grossly under utilized with the exceptions of TV bands 470 - 600 MHz and 730 - 860 MHz which have a duty cycle of 29.52% and 15.02% respectively. GSM 1800 band with a duty cycle of 51.57% appear to be much more utilized than GSM 900 band which has an estimated duty cycle of 40.85%. From the results of the survey, GSM 1800 band has the highest duty cycle. The duty cycle of GSM 900 and GSM 1800

bands shown in subplot 3 of Figures 3.11 and 3.12 exhibit a different trend with no frequency resource block having a duty cycle of zero, contrary to those of Figures 3.5 to 3.10 where there are many frequency resource blocks (spectrum) with zero duty cycle for the whole period of measurements (observations). Based on the studied location, the ISM 2.4 GHz band has a higher spectrum occupancy with a duty cycle of 18.55% than ISM 868 MHz which has a duty cycle of 14.71%. From the aforementioned, it is obvious that over 50% of the licensed spectrum are not utilized. Thus, the current spectrum scarcity is a direct consequence of static spectrum allocation policies and not the fundamental lack of spectrum.

Chapter 4 detailed the implementation of the RF power prediction models, and the topologies adopted. The results of the sensitivity analysis shows that the first four most recent past RF power samples account for 94.68% of the total decrease in MSE resulting from all the past RF power samples used. Hence in this research, 4 past recent RF power samples were used as part of the ANN and SVM inputs for current RF power prediction.

A combine RF power prediction model consists of six prediction models i.e. three ANNs topologies (FFB, CFF and LR) and three SVM regression models using linear, RBF and exponential kernel was designed. The weights of the combine model were evolved using the proposed method presented in Sections 4.4 and 4.4.1, and the accuracy of the combine model increases with an average of 46% as compare with the individual models used.

To minimised the problem of premature convergence which is one of the major setback of back propagation algorithm, and to have a robust ANN model that is well generalised, the ANNs were trained using a hybrid algorithm consisting of different variants of population based and compact metaheuristic optimization algorithms, and back propagation algorithm (BPA). The hybrid algorithm exploit the global search advantages of population based randomization optimization algorithms and the local search capability of single solution BPA. The initial weights were evolved using differential evolution (DE) and particle swarm optimization (PSO) variants as detailed in Section 2.6, after which the weights are fine tuned toward the global optimum using BPA. The models trained with the hybrid algorithm was found to outperform the ones train with only BPA with a MSE of 0.0242 while BPA have a MSE of 0.1345, Iliya et al. (2015a;d; 2014b).

The SVM is trained using a hybrid ad-hoc nested algorithm whose outer layer consists of different variants of population based and compact metaheuristic optimization algorithms for sampling the SVM meta-parameters (C parameter, ε and the kernel parameter γ), while the inner layer is a convex optimization algorithm for evolving the weights and bias of the SVM.

For the models designed for prediction of RF power in one channel, the feed forward ANN outperformed the other models with an average mean square error (AMSE) of 0.0215. The best SVM model is obtained using exponential kernel with an AMSE

of 0.0226. For the multiple channel, single output prediction models, the AMSE of feed forward ANN is 0.0565 while that of Linear kernel SVM is 0.0640. Base on the results obtained, both ANN and SVM are good choices for this problem even though the ANN outperformed the SVM in all cases. Considering the computing demand, and the memory and power constrain of cognitive radios, the recommended model is the feed forward ANN model designed for prediction of RF power in multiple channels but one at a time. From the results presented in Tables 4.2 and 4.3 with regard to this problem, the differential evolution algorithms are preferred to the swarm intelligence algorithms for training the models.

Spectrum holes prediction using probabilistic ANN model and the prediction using two algorithmic methods (i.e. blind linear and blind stochastic search methods) are compared. The word probabilistic here referred to the fact that the outputs of the ANN gives the probability of the channels been idle. In this simulation, it was assumed that all the channels have the same noise and quality of service, and only one channel is needed for communication. The spectrum activities of 20 channels were modelled for different traffic intensities as described in Section 4.5.2. Each method is evaluated based on the number of channels sensed out of the 20 channels before the first idle channel is found. For each search, the ANN ranked the channels in decreasing order of being idle and start searching from the channel with the highest probability of being idle, the search stop when the first idle channel is found. The blind linear method always start searching from the first channel serially while the blind stochastic method searches randomly. For traffic intensity of 32.62%, out of 7440 search times, 97.87% (7281) of the total search the ANN found spectrum hole at first search while that of the blind linear and blind stochastic search are 62.88% and 65.44% respectively as shown in Figure 4.14. For all the traffic intensities considered in this study, the ANN outperformed the other two algorithmic approaches as shown in Table 4.6. The probabilistic ANN used in this research also outperformed the ANN used by Baldo et al. (2009) for dynamic channel selection.

Chapter 5 consist of one of the major contribution of this thesis. ANN and SVM models were used for segmenting short impaired speech utterance into silent/noise, unvoiced, and voiced sections. Moreover, the proposed technique identifies those points of the impaired speech where the spectrum becomes steady. The resulting technique thus aims at detecting that limited section of the speech which contains the information about the potential impairment. This section is of interest to the speech therapist as it is likely to contain the information about the speech organs (articulators) responsible for the impairment, Iliya and Neri (2016), Iliya et al. (2014d;c). Five speech features consisting of both temporal (time) and spectral (frequency) features were extracted from the speech signal as described in Section 5.4. These five feature are used as inputs for training the models for identification of the three segments of the impaired speech. When the models were tested with impaired speech from both male and female subjects, the SVM segmentation models outperformed the ANN models with percentage error (PE) of 1.28%, 1.24% and 1.95% while the PE for the ANN models are 1.57%,

1.48% and 2.20% for detection of silence/noise, unvoiced and steady state segments respectively. In the proposed disordered speech training system, the patient is requested to speak only short utterances of the form vowel - consonant - vowel e.g. ASA. The leading vowel sound is added to make it easier for the patient to pronounce the following part. The steady state of interest is the segment within the second vowel where the spectrum become steady in the frequency domain (i.e. the formants frequencies remain constant). Within the steady state, there is no new spectral information. Hence the most informative section of the impaired speech is likely to be at the beginning of the unvoiced speech (consonant) and the first frame within the steady state. The proposed training system, efficiently captured the informative segment of the impaired speech which can be used for further analysis towards the realization of a standalone artificial speech therapist.

In addition to this, a robust user friendly GUI tool was developed using matlab for normal and impaired speech signal processing that contains both algorithmic signal processing methods and computation intelligence framework approaches as presented in Section 5.4.7. The recordings used for the training and testing of the models presented in this thesis, were done in English language, to increase the robustness of the tool and to cover a wide range of languages and voice types, the tool developed allow the user to train the model using any language of their choice. The tool can also be used for visualization of many speech features and to observe their dynamics as related to the various sections of the impaired speech. This is very important as it enable the choice of the most relevant features for training and prediction.

6.3 Future Work

Spectrum Occupancy Survey: The dataset used in this research was captured within one fixed geographical location and at the same height for the entire period of the experiment. This is not sufficient to provide a robust and generalised information about the spectrum occupancy estimate of the studied environment. To provide a wider view of spectrum occupancy estimate that can facilitate cognitive radio network implementation, spectrum measurements campaigns should be conducted simultaneously in more than one geographical location at different heights within the same period of time. The collation and analysis of the data obtained from the independent measurements will give a more accurate estimate that can be used to identify underutilised bands with reference to frequency, time and spatial domains.

Spectrum Hole Prediction: One of the initial objectives of this research is to develop CI predictive models using historic data to predict in real time the current channels statute i.e. either vacant (idle) of busy using a binary classifier. This will enable the CR to predictively select idle channel (spectrum hole) for communication. But after capturing the historic data for training the models, the actual channels statutes (busy or idle) can only be known if the characteristics of the signal such as modulation type, encoding schemes, etc is known. Since the characteristics of the signals captured

are not known, the actual channels statute could not be determined. As a result, all the prediction models implemented in this research can only be used to predict real world RF power of the channels not the actual channels statute. This will help the cognitive radio to avoid noisy channels, and if given the right threshold, the actual channel statute can be known. That is, if the predicted power is above the threshold, the channel is busy otherwise it is idle. The models implemented for prediction of channels statute (spectrum holes) presented in Section 4.5.2 were designed using simulated data. Hence spectrum sensing in search of spectrum holes without a priory knowledge of the primary users signal type, is currently an active research field.

Realization of Standalone Artificial Speech Therapist: Having been able to capture the parts of the impaired speech that are likely to contain the information about the potential impairment, future work should aim at using other high dimensional temporal (time) and frequency domain speech features along with other CI models, to analysis in detail the informative sections of the impaired speech, and to identify in time or near real time the speech organs (articulators) that are responsible for each segments of the impaired speech, and possible correction measures. This will be followed by The final phase which will focus on the development of a robust artificial speech therapist that can serve as a self training system for speech disordered patient with reliable feedback progress report in form of text and audiovisuals.

Bibliography

A. O. Abioye, A. Kola-Mustapha, G. T. Chi, and S. Iliya. Quantification of in situ granulation-induced changes in pre-compression, solubility, dose distribution and intrinsic in vitro release characteristics of ibuprofen–cationic dextran conjugate crystanules. *International Journal of Pharmaceutics*, (471):453–477, 2014.

Y. Adane, M. Bavaro, G. John, G. Eric, and M. Dumville. Miniaturized dual-channel high-end receivers for gnss software defined radio applications. December 2008.

Y. Adane, M. Bavaro, M. Dumville, G. Eric, and G. John. Low cost multi constellation front end for gnss software defined receivers. In *European Navigation Conference - Global Navigation Satellite Systems*, 2009.

C. W. Ahn and R. Ramakrishna. Elitism-based compact genetic algorithms. *Evolutionary Computation, IEEE Transactions on*, 7(4):367–385, Aug 2003. ISSN 1089-778X. doi: 10.1109/TEVC.2003.814633.

K. R. Aida–Zade, C. Ardil, and S. Rustamov. Investigation of combined use of mfcc and lpc features in speech recognition systems. *International Journal of Computer, Information Science and Engineering*, 2, 2008.

A. Akutekwe, S. Iliya, and H. Seker. An optimized hybrid dynamic bayesian network approach using differential evolution algorithm for hepatocellular carcinoma diagnosis. In *IEEE International Conference on Adaptive Science and Technology (ICAST)*, 2014.

M. Almaraashi, R. John, S. Coupland, and A. Hopgood. Time series forecasting using a tsk fuzzy system tuned with simulated annealing. In *Proceedings of FUZZ-IEEE2010, World Congress on Computational Intelligence*, 2010.

E. Alpaydin. *Introduction to Machine Learning*. Adaptive computation and machine learning. MIT Press, 2004. ISBN 9780262012119.

N. Arora and R. Mahajan. Cooperative spectrum sensing using hard decision fusion scheme. *International Journal of Engineering Research and General Science*, 2, June 2014.

R. Azmi. Support vector machine based white space predictors for cognitive radio. Master's thesis, 2011.

R. Bacchus, T. Taher, K. Zdunek, and D. Roberson. Spectrum utilization study in support of dynamic spectrum access for public safety. In *New Frontiers in Dynamic Spectrum, 2010 IEEE Symposium on*, pages 1–11, April 2010. doi: 10.1109/DYSPAN.2010.5457871.

N. Baldo, B. R. Tamma, B. Manoj, R. Rao, and M. Zorzi. A neural network based cognitive controller for dynamic channel selection. In *IEEE Communications Society, IEEE ICC 2009 proceedings*, 2009.

S. Bernhard and S. A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001. ISBN 0262194759.

P. P. Bhattacharya, R. Khandelwal, R. Gera, and A. Agarwal. Smart radio spectrum management for cognitive radio. *International Journal of Distributed and Parallel Systems (IJDPS)*, 2(4), July 2011.

C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science & Business Media, LLC, 233 Spring Street, New York, NY 10013, USA, February 2006.

J. Brest, S. Greiner, B. Bošković, M. Mernik, and V. Žumer. Self-Adapting Control Parameters in Differential Evolution: A Comparative Study on Numerical Benchmark Problems. *IEEE Transactions on Evolutionary Computation*, 10(6):646–657, 2006.

S. Brognaux and D. T. Hmm-based speech segmentation: Improvements of fully automatic approaches. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 24(1):5–15, 2016.

R. Bureau. "handbook spectrum monitoring. *International Telecommunication Union (ITU)*, 2002.

A. Caponio, F. Neri, and V. Tirronen. Super-fit control adaptation in memetic differential evolution frameworks. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 13:811–831, 2009.

F. Caraffini, F. Neri, G. Iacca, and A. Mol. Parallel memetic structures. *Information Sciences*, 227(0):60 – 82, 2013. ISSN 0020-0255.

F. Caraffini, F. Neri, B. Passow, and G. Iacca. Re-sampled inheritance search: High performance despite the simplicity. *Soft Computing*, 17(12):2235–2256, 2014a.

F. Caraffini, F. Neri, and L. Picinali. An analysis on separability for memetic computing automatic design. *Information Sciences*, 265:1–22, 2014b.

Y. Chen and H.-S. Oh. A survey of measurement-based spectrum occupancy modelling for cognitive radios. In *1553-877X ©2013 IEEE IEEE Communications Surveys and Tutorials*, 2013. doi: 10.1109/COMST.2014.2364316.

V. Cherkassky and Y. Ma. Practical selection of svm parameters and noise estimation for svm regression. *Neural Networks*, 17:113–126, 2004.

V. S. Cherkassky and F. Mulier. *Learning from Data: Concepts, Theory, and Methods*. John Wiley and Sons, Inc., New York, NY, USA, 1st edition, 1998. ISBN 0471154938.

S. Chu, S. Narayanan, and C. C. J. Kuo. Environmental sound recognition with time–frequency audio features. *IEEE Transactions on Audio, Speech, and Language Pprocessing*, 17(6), August 2009.

Communication. Federal comminucation commission notice of inquiry and notice of proposed rule making, in the matter of establishment of an interference temperature metric to quantify and manage interference and to expand available unlicensed operation in certain fixed, mobile and satellite frequency bands. (03-237), November 13, 2003.

S. Contreras, G. Villardi, R. Funada, and H. Harada. An investigation into the spectrum occupancy in japan in the context of tv white space systems. In *Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), 2011 Sixth International ICST Conference on*, pages 341–345, June 2011a.

S. Contreras, G. Villardi, R. Funada, and H. Harada. An investigation into the spectrum occupancy in japan in the context of tv white space systems. In *6th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, 2011b.

P. Cornelius, N. Higgett, and R. Kaleem. Sensory articulation speech system: Sassy - a 3d animation based therapeutic application for motor speech disorders. *Journal of Assistive Technologies*, 5(3):123–130, 2011.

Y. Couce, L. Franco, D. Urda, J. Subirats, and M. J. Jose. Hybrid generalization-correlation method for feature selection in high dimensional dna microarray prediction problems. In *Proceedings of the IWANN 2011 conference, Torremolinos, Spain*, pages 8–10, June 2011.

K. Dash, D. Padhi, B. Panda, and S. Mohanty. Speaker identification using mel frequency cepstralcoefficient and bpnn. *International Journal of Advanced Research in Computer Science and Software Engineering Research Paper*, 2, April 2012.

D. Datla. Spectrum surveying for dynamic spectrum access networks. Msc thesis, University of Madras, India, 2004.

G. E. H. David E. Rumelhart and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

M. Davide and H. Simon. Advances in kernel methods. chapter Support Vector Machines for Dynamic Reconstruction of a Chaotic System, pages 211–241. MIT Press, Cambridge, MA, USA, 1999. ISBN 0-262-19416-3.

H. Demuth and M. Beale. *Neural Network Toolbox for Use with MATLAB*. July 2002.

N. S. Dey, R. Mohanty, and K. L. chugh. Speech and speaker recognition system using artificial neural networks and hidden markov model. In *978-0-7695-4692-6/12 $26.00 ©2012 IEEE International Conference on Communication Systems and Network Technologies*, 2012.

A. E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing*, pages 175–188. Springer Verlag, Berlin, 2003.

H. Ekram, N. Dusit, and Z. Han. *Dynamic Spectrum Access and Management in Cognitive Radio Networks*. Cambridge University Press, New York, NY, USA, 1st edition, 2009. ISBN 0521898471, 9780521898478.

Z. Fang, C. S. Shing, C. Ching, H. Yihong, and W. Zhaoan. Use hamming window for detection the harmonic current based on instantaneous reactive power theory. In *Power Electronics and Motion Control Conference, 2004. IPEMC 2004. The 4th International*, volume 2, pages 456–461 Vol.2, Aug 2004.

D. Fogel. Review of computational intelligence: imitating life. *IEEE Transactions on Neural Networks*, 6:1562 – 1565, 1995.

L. Franco and S. Cannas. Generalization and selection of examples in feed-forward neural networks. *Neural Computation*, pages 2405 – 2426, 2000.

R. Frank. The perceptron: a probabilistic model for information storage and organisation in the brain. *Psychology Review*, 65:386–408, 1958.

W. Gardner. Exploitation of spectral redundancy in cyclostationary signals. *Signal Processing Magazine, IEEE*, 8(2):14–36, April 1991. ISSN 1053-5888. doi: 10.1109/79.81007.

A. Garhwal and P. P. Bhattacharya. A survey on spectrum sensing techniques in cognitive radio. *International Journal of Computer Science and Communication Networks*, 2:196–206, Nov 2011.

S. Geirhofer, J. Z. Sun, L. Tong, and B. M. Sadler. Cognitive frequency hopping based on interference prediction: Theory and experimental results. 13(2), march 17 2009.

L. Giupponi and A. Perez. Fuzzy-based spectrum handoff in cognitive radio networks. 2008.

I. Gómez, S. A. Cannas, J. J. Omar Osenda, and L. Franco1I. The generalization complexity measure for continuous input data. *The Scientific World Journal*, 2014. doi: http://dx.doi.org/10.1155/2014/815156.

M. A. Gongora and D. Elizondo. Current trends on knowledge extraction and neural networks. In *ICANN 2005 on Neural Networks and Knowledge Extraction*, 2005. doi: http://dx.doi.org/10.1007/11550907.

G. Harik, F. Lobo, and D. Goldberg. The compact genetic algorithm. *Evolutionary Computation, IEEE Transactions on*, 3(4):287–297, Nov 1999. ISSN 1089-778X. doi: 10.1109/4235.797971.

S. Haykin. *Neural Networks and Learning Machines*. Pearson Education, Inc., Upper Saddle River, New Jersey 07458, 3rd edition, 2008.

S. Haykin, D. J. Thomson, and J. H. Reed. Spectrum sensing for cognitive radio. In *IEEE Transactions on Cognitive Radio*, May 2009.

S. Hiremath and S. K. Patra. Transmission rate prediction for cognitive radio using adaptive neural fuzzy inference system. In *IEEE 5th International Conference on Industrial and Information Systems (ICIIS), India*, Aug 2010.

D. Horgan and C. C. Murphy. Voting rule optimisation for double threshold energy detector-based cognitive radio networks. In *Signal Processing and Communication Systems (ICSPCS), 2010 4th International Conference*, pages 1–8, Dec 2010. doi: 10.1109/ICSPCS.2010.5709679.

A. Huemer, G. Mario, and E. David. Evolving a neural network using dyadic connections. In *Proceedings: International Joint Conference on Neural Networks (IJCNN 2008)*, pages 1019–1025, 2008a. ISBN 9781424418206. doi: http://dx.doi.org/10.1109/IJCNN.2008.4633924.

A. Huemer, G. Mario, and D. Elizondo. Self constructing neural network robot controller based on on-line task performance feedback. In *Proceedings: 5th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, pages 326–333, 2008b. ISBN 9789898111302.

A. Huemer, D. Elizondo, and M. Gongora. *A constructive neural network for evolving a machine controller in real-time, Book chapter in: Constructive Neural Networks*, volume 258. Springer Berlin Heidelberg, 2009. doi: http://dx.doi.org/10.1007/978-3-642-04512-7%_12.

M. Huemer, A. Gongora and D. Elizondo. A robust reinforcement based self constructing neural network. In *The 2010 International Joint Neural Networks (IJCNN)*, 2010. doi: http://dx.doi.org/10.1109/IJCNN.2010.5596762.

G. Iacca, R. Mallipeddi, E. Mininno, F. Neri, and P. N. Suganthan. Super-fit and Population Size Reduction Mechanisms in Compact Differential Evolution. In *Proceedings of IEEE Symposium on Memetic Computing*, pages 21–28, 2011a.

G. Iacca, E. Mininno, and F. Neri. Composed compact differential evolution. *Evolutionary Intelligence*, 4(1):17–29, 2011b.

G. Iacca, F. Caraffini, E. Mininno, and F. Neri. Robot base disturbance optimization with compact differential evolution light. In *EvoApplications*, Lecture Notes in Computer Science, pages 285–294. Springer, 2012a.

G. Iacca, F. Caraffini, and F. Neri. Compact differential evolution light. *Journal of Computer Science and Technology*, 27(5):1056–1076, 2012b.

S. Iliya and F. Neri. Towards artificial speech therapy: A neural system for impaired speech segmentation. *International Journal of Neural Systems (IJNS)*, 26(6), March 2016. doi: 10.1142/S0129065716500234.

S. Iliya, E. Goodyer, M. Gongora, J. Gow, and J. Shell. Optimized artificial neural network using differential evolution for prediction of rf power in vhf/uhf tv and gsm 900 bands for cognitive radio networks. In *978-1-4799-5538-1/14$31.00 ©2014 IEEE International Conference on Computational Intelligence, UKCI*, 2014a.

S. Iliya, E. Goodyer, J. Shell, J. Gow, and M. Gongora. Optimized neural network using differential evolutionary and swarm intelligence optimization algorithms for rf power prediction in cognitive radio network: A comparative study. In *978-1-4799-4998-4/14/$31.00 ©2014 IEEE International Conference on Adaptive Science and Information Technology*, 2014b.

S. Iliya, D. Menzies, F. Neri, P. Cornelius, and L. Picinali. Robust impaired speech segmentation using neural network mixture model. In *2014 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 444–449, Dec 2014c. doi: 10.1109/ISSPIT.2014.7300630.

S. Iliya, F. Neri, D. Menzies, P. Cornelius, and L. Picinali. Differential evolution schemes for speech segmentation: A comparative study. In *IEEE Symposia Series on Computational Intelligence (SSCI)*, 2014d.

S. Iliya, Eric, J. Gow, J. Shell, and M. Gongora. Application of artificial neural network and support vector regression in cognitive radio networks for rf power prediction using compact differential evolution algorithm. In *Proceedings of Federated Conference on Computer Science and Information Systems (FedCSIS), 2015*, pages 55–66, Sept 2015a. doi: 10.15439/2015F14.

S. Iliya, E. Goodyer, M. Gongora, and J. Gow. Spectrum hole prediction and white space ranking for cognitive radio network using an artificial neural network. *International Journal of Scientific and Technology Research*, 4, August 2015b. ISSN 2277-8616.

S. Iliya, E. Goodyer, J. Gow, M. Gongora, and J. Shell. Spectrum occupancy survey in leicester, uk, for cognitive radio application. *International Journal of Scientific and Engineering Research*, August 2015c.

S. Iliya, E. Goodyer, J. Gow, J. Shell, and M. Gongora. Application of artificial neural network and support vector regression in cognitive radio networks for rf power prediction using compact differential evolution algorithm. *Annals of Computer Science and Information Systems*, 5:55–66, Sept 2015d. doi: 10.15439/2015F14.

M. H. Islam, C. L. Koh, S. W. Oh, X. Qing, Y. Y. Lai, C. Wang, Y. C. Liang, B. E. Toh, F. Chin, G. L. Tan, and W. Toh. Spectrum survey in singapore: Occupancy measurements and analyses. In *Proc. CrownCom 2008*, pages 1 – 7, May 2008.

T. K. James and I. W. Tsang. Linear dependency between epsilon and the input noise in epsilon-support vector regression. In *IEEE Transactions on Neural Networks*, 2003.

J. Jerez, L. Franco, E. Alba, A. Llombart-Cussac, A. Lluch, N. Ribelles, B. Munarriz, and M. Martın. Improvement of breast cancer relapse prediction in high risk intervals using artificial neural networks. *Breast Cancer Research and Treatment*, pages 265–272, 2005.

J. Jerez, I. Molina, P. G. Laencina, E. Alba, N. Ribelles, M. Martın, and L. Franco. Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial Intelligence in Medicine*, pages 105–115, 2010.

S. Jethro, C. Samuel, and G. Eric. Fuzzy data fusion for fault detection in wireless sensor networks. In *2010 UK Workshop on Computational Intelligence, UKCI 2010*, 2010. doi: http://dx.doi.org/10.1109/UKCI.2010.562559. URL http://hdl.handle.net/2086/4492.

Z. Jianli. Based on neural network spectrum prediction of cognitive radio. In *978-1-4577-0321-8/11/$26.00 ©2011 IEEE*, 2011.

S.-T. John and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, July 2004.

K. A. D. Jong. *Evolutionary Computation A Unified Approach*. The MIT Press Cambridge, Massachusetts, London, England, 2006.

N. Kamil and X. Yuan. Detection proposal schemes for spectrum sensing in cognitive radio. *Wireless Sensor Network*, 2:365–372, 2010. doi: 10.4236/wsn.2010.24048.

V. Kecman. *Learning and soft computing*. MIT Press Cambridge, Mass, 2001.

R. Lawrence and R. B. Gold. *Theory and Application of Digital Signal Processing*. Prentice-Hall, 1975.

W. Y. Lee. *spectrum management in cognitive radio wireless networks*. PhD thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology, August 2009.

D. Li and O. Douglas. *Speech Processing: A Dynamic and Optimization-Oriented Approach (Signal Processing and Communications)*. Marcel Dekker Inc., New York, NY, U.S.A, 2003.

D. Li, L. Xu, E. D. Goodman, Y. Xu, and Y. Wu. Integrating a statistical background-foreground extraction algorithm and SVM classifier for pedestrian detection and tracking. *Integrated Computer-Aided Engineering*, 20(3):201–216, 2013.

X. Li and S. A. Zekavat. Traffic pattern prediction and performance investigation for cognitive radio systems. In *IEEE Communication Society, WCNC Proceedings*, 2008.

J. J. Liang, A. K. Qin, P. N. Suganthan, and S. Baskar. Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. *IEEE Transactions on Evolutionary Computation*, 10(3):281–295, 2006.

C. P. Lim, S. C. Woo, A. S. Loh, and R. Osman. Speech recognition using artificial neural networks. In *Proceedings of the IEEE Conference on Web Information Systems Engineering*, 2000.

Z. Lin, X. Jian, L. Huang, and Y. Yao. Energy prediction based spectrum sensing approach for cognitive radio network. In *978-1-4244-3693-4/09/$25.00 ©2009 IEEE*, 2009.

R. M. Luque-Baena, D. Urda1, J. L. Subirats, L. Franco1, and J. M. Jerez. Application of genetic algorithms and constructive neural networks for the analysis of microarray cancer data. *Theoretical Biology and Medical Modelling*, 2014.

M. Matinmikko, J. Del Ser, T. Rauma, and M. Mustonen. Fuzzy-logic based framework for spectrum availability assessment in cognitive radio systems. *Selected Areas in Communications, IEEE Journal on*, 31(11):2173–2184, November 2013. ISSN 0733-8716. doi: 10.1109/JSAC.2013.131117.

S. Matthews, M. Gongora, and A. Hopgood. Evolutionary algorithms and fuzzy sets for discovering temporal rules. *International Journal of Applied Mathematics and Computer Science*.

S. Matthews, M. Gongora, and A. Hopgood. Evolving temporal fuzzy itemsets from quantitative data with a multi-objective evolutionary algorithm. In *IEEE 5th Int. Workshop on Genetic and Evolutionary Fuzzy Systems*, 2011.

M. A. McHenry, P. A. Tenhula, and D. McCloskey. Chicago spectrum occupancy measurements & analysis and a long-term studies proposal. In *Proc. TAPAS 2006*, 2006.

M. Mehdawi, N. Riley, K. Paulson, A. Fanan, and M. Ammar. Spectrum occupancy survey in hull-uk for cognitive radio applications: Measurement & analysis, April 2013.

E. Mininno, F. Neri, F. Cupertino, and D. Naso. Compact Differential Evolution. *IEEE Transactions on Evolutionary Computation*, 15(1):32–54, 2011a.

E. Mininno, F. Neri, F. Cupertino, and D. Naso. Compact differential evolution. *Evolutionary Computation, IEEE Transactions on*, 15(1):32–54, Feb 2011b. ISSN 1089-778X. doi: 10.1109/TEVC.2010.2058120.

J. Mitola and G. Maguire. Cognitive radio: making software radios more personal. *IEEE Personal Communications*, 6(4):13–18, Aug 1999. ISSN 1070-9916. doi: 10.1109/98.788210.

J. Montgomery. Crossover and the different faces of differential evolution searches. In *IEEE Congress on Evolutionary Computation*, pages 1–8, July 2010. doi: 10.1109/CEC.2010.5586184.

F. Neri and V. Tirronen. Recent Advances in Differential Evolution: A Review and Experimental Analysis. *Artificial Intelligence Review*, 33(1–2):61–106, 2010.

OFCOM. Cognitive radio technology: A study for ofcom summary report. 1, Febuary 2007.

J. Oh and W. Choi. A hybrid cognitive radio system: A combination of underlay and overlay approaches. In *IEEE Transactions on Cognitive Radio*, 2009.

F. Ortega-Zamorano, J. M. Jerez, D. Urda, R. M. Luque-Baena, and L. Franco. Efficient implementation of the backpropagation algorithm in fpgas and microcontrollers. *IEEE Transactions on Neural Networks and Learning Systems*, 2015. doi: http://dx.doi.org/10.1109/TNNLS.2015.2460991.

K. Padmanabhan. *A Practical Approach to Digital Signal Processing*. New Age International Publishers, 2003.

K. Patil, K. Skouby, A. Chandra, and R. Prasad. Spectrum occupancy statistics in the context of cognitive radio. In *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, pages 1–5, Oct 2011.

C. Penaloza, I. Heikius, and J. Sonja. Exploring speech segmentation abilities in people with chronic aphasia. *Procedia - Social and Behavioral Sciences*, 94:112 – 115, 2013.

C. Penaloza, A. Benetello, L. Tuomiranta, I.-M. Heikius, and J. Sonja. Speech segmentation in aphasia. *Aphasiology*, 29:724 – 743, 2015.

H. Perez-Meana. *Advances in Audio and Speech Signal Processing: Technologies and Applications*. 701 E. Chocolate Avenue, Hershey PA 17033, U.S.A, 2007.

L. Perez-Rivas, J. Jerez, C. F. de Sousa, V. de Luque, C. Quero, B. Pajares, L. Franco, N. R. Alfonso Sanchez-Muñoz, and E. Alba. Serum protein levels following surgery in breast cancer patients: A protein microarray approach. *International Journal of Oncology*, pages 2200–2206, 2012.

P. Podder, T. Z. Khan, M. H. Khan, and M. Rahman. Comparative performance analysis of hamming, hanning and blackman window. *International Journal of Computer Application*, 96(18), June 2014.

K. V. Price, R. Storn, and J. Lampinen. *Differential Evolution: A Practical Approach to Global Optimization*. Springer, 2005.

P. Prithvi, A. Kumar, and T. K. Kumar. Speaker independent speech recognition of english digits using hybridized vq-hmm model in noisy environment. *International Journal of Engineering Research & Technology (IJERT)*, 3, April 2014.

J. Proakis and M. Salehi. *Digital Communications*. Mcgraw-Hill, 5th edition, 2008.

K. A. Qaraqe, H. Celebi, A. Gorcin, A. El-Saigh, H. Arslan, and M. S. Alouini. Empirical results for wideband multidimensional spectrum usage. In *Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*, pages 1262–1266, Sept 2009. doi: 10.1109/PIMRC.2009.5450192.

F. QI, C. BAO, and Y. LIU. A novel two-step svm classifier for voiced/unvoiced/silence classification of speech. 2004.

A. K. Qin, V. L. Huang, and P. N. Suganthan. Differential evolution algorithm with strategy adaptation for global numerical optimization. In *IEEE Transactions on Evolutionary Computation*, volume 13, April 2009.

M. Radmard, M. Hadavi, and M. M. Nayebi. A new method of voiced/unvoiced classification based on clustering. *Journal of Signal and Information Processing*, pages 336–347, 2011.

N. Ribelles, L. Perez-Villa, J. M. Jerez, B. Pajares, L. Vicioso, B. Jimenez, V. de Luque, L. Franco, E. Gallego, A. Marquez, M. Alvarez, A. Sanchez-Munoz, L. Perez-Rivas1, and E. Alba. Pattern of recurrence of early breast cancer is different according to intrinsic subtype and proliferation index. *Breast Cancer Research*, 2013.

T. W. Rondeau, B. Le, C. J. Rieser, and C. W. Bostian. Cognitive radios with genetic algorithms: Intelligent control of software defined radios. In *©2004 SDR Forum, Proceeding of the SDR 2004 Technical Conference and Product Exposition*, 2004.

G. Saha, S. Chakroborty, and S. Senapati. A new silence removal and endpoint detection algorithm for speech and speaker recognition applications. In T. S. Lamba, P. K. Biswas, and S. S. Pathak, editors, *Proceedings of the Eleventh National Conference on Communications*, 2005.

M. Sahidullah and G. Saha. Design analysis and experimental evaluation of block based transformation in mfcc computation for speaker recognition. *Speech Communication*, 54(4):543 – 565, 2012.

S. Sharma. Cognitive radio approach : Spectrum sensing. *Global Journal of Researches in Engineering*, 13, 2013.

Y. Shi and R. Eberhart. A modified particle swarm optimizer. In *Proceedings of the IEEE Congress on Evolutionary Computation*, 1998.

N. Siddique and H. Adeli. *Computational Intelligence Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing*. John Wiley & Sons, Ltd., Southern Gate, Chichester, West Sussex, PO19 8SQ, UK, 2013.

S. M. Siniscalchi, D. Yu, L. Deng, and C.-H. Lee. Speech recognition using long-span temporal patterns in a deep network model. In *IEEE Signal Processing Letter*, volume 20, March 2013.

A. J. Smola, N. Murata, B. Scholkopf, and K. R. Muller. Asymptotically optimal choice of ε- loss for support vector machines. In *Proceedings of the International Conference on Artificial Neural Networks*, pages 105–110. Springer, Berlin, 1998.

C. Stevenson, G. Chouinard, Z. Lei, W. Hu, J. Stephen, and W. Caldwell. The first cognitive radio wireless regional area network standard. In *IEEE 802.22*, 2009.

R. Storn and K. Price. Differential Evolution - a Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, 11 (TR-95-012):341–359, 1997.

M. Subhedar and G. Birajdar. Spectrum sensing techniques in cognitive radio network: A survey. *International Journal of Next-Generation Networks (IJNGN)*, 3(2), June 2011a.

M. Subhedar and G. Birajdar. Spectrum sensing techniques in cognitive radio networks: A survey. *International Journal of Next-Generation Networks*, 3(2):37–51, 2011b.

J. L. Subirats, J. Jerez, I. Gomez, and L. Franco. Multiclass pattern recognition extension for the new c-mantec constructive neural network algorithm. *Cognitive Computation*, pages 285–290, 2010.

J. Sun, H. Li, and H. Adeli. Concept drift-oriented adaptive and dynamic support vector machine ensemble with time window in corporate financial risk prediction. *IEEE T. Systems, Man, and Cybernetics: Systems*, 43(4):801–813, 2013.

A. H. Sung. Ranking importance of input parameters of neural networks. *Expert Systems with Applications*, 15(3):405–411, November 1998.

K. Suzuki. *Artificial Neural Networks: Industrial and Control Engineering Applications*. InTech, India, 1 edition, 2011.

Z. Tabakovic, S. Grgic, and M. Grgic. Fuzzy logic power control in cognitive radio. In *IEEE transactions*, 2009.

M. I. Taj and M. Akil. Cognitive radio spectrum evolution prediction using artificial neural networks based mutivariate time series modelling. In *European Wireless, Vienna Austria*, April 2011.

S. K. Udgata, K. P. Kumar, and S. L. Sabat. Swarm intelligence based resource allocation algorithm for cognitive radio network. In *Parallel Distributed and Grid Computing (PDGC), 2010 1st International Conference on*, pages 324–329, Oct 2010. doi: 10.1109/PDGC.2010.5679975.

V. Valenta, R. Marsalet, G. Baudoin, M. Villegas, M. Suarez, and F. Robert. Survey on spectrum utilization in europe: Measurements, analyses and observations. In *Cognitive Radio Oriented Wireless Networks Communications (CROWNCOM), 2010 Proceedings of the Fifth International Conference on*, pages 1–5, June 2010.

V. Vapnik. *Statistical learning theory.* New York: Wiley, 1998.

V. Vapnik. *The nature of statistical learning theory.* Springer-Verlag New York Inc, 1999.

C. Vladimir and Y. MA. Selection of meta-parameters for support vector regression. pages 687–693, August 2002.

Z. Wang. *Measurement and Modelling of Spectrum Occupancy.* Phd thesis, Ustinov College, University of D u r h a m, India, 2009.

M. Warren and P. Walter. A logical calculus of the ideas imminent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

W. Wenjian, Z. Xu, W. Lu, and X. Zhang. Determination of the spread parameter in the gaussian kernel for classification and regression. 55(3):643–663, October 2003.

F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6): 80–83, 1945.

O. Winston, A. Thomas, and W. Okelloodongo. Optimizing neural network for tv idle channel prediction in cognitive radio using particle swarm optimization. In *Computational Intelligence, Communication Systems and Networks (CICSyN), 2013 Fifth International Conference on*, pages 25–29, June 2013. doi: 10.1109/CICSYN.2013.68.

M. Wolfel and J. McDonough. *Distant Speech Recognition.* John Wiley, 2009.

S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. 1:356–366, 1932.

A. M. Wyglinski, M. Nekovee, and Y. T. Hou. *Cognitive Radio Communications and Networks.* 2009.

A. W. Wyglinski, M. Nekovee, and Y. T. Hou. *Cognitive Radio Communications and Networks Principles and Practice.* Elsevier, 2010.

T. Xiaoyun, T. pengwu, and Y. hongyi. Modulation classification based on spectral correlation and svm. In *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on*, pages 1–4, Oct 2008. doi: 10.1109/WiCom.2008.409.

X. Xing, T. Jing, W. Cheng, Y. Huo, and X. Cheng. Spectrum prediction in cognitive radio networks. In *1536-1284/13/$25.00 ©2013 IEEE Transactions on Wireless Communications*, April 2013.

X. Yin-shan, M. Ya-dong, and Y. Ting. Combined forecasting model of urban water demand under changing environment. In *978-1-4577-0290-7/11/$26.00 ©2011 IEEE*, 2011.

T. Yucek and H. Arslan. A survey of spectrum sensing algorithms for cognitive radio applications. *Communications Surveys Tutorials, IEEE*, 11(1):116–130, First 2009. ISSN 1553-877X. doi: 10.1109/SURV.2009.090109.

L. Zadeh. Roles of soft computing and fuzzy logic in the conception, design and deployment of information/intelligent systems. 1998.

D. Zaharie. A comparative analysis of crossover variants in differential evolution. In *Proceedings of the International Multiconference on Computer Science and Information Technology*, pages 171–181, 2007.

S. Zarrin. *Spectrum Sensing in Cognitive Radio Network*. PhD thesis, Department of Electrical and Computer Engineering, University of Toronto, 2011.